DEVELOPMENT OF AN ACOUSTIC CLASSIFICATION SYSTEM FOR PREDICTING ROCK STRUCTURAL STABILITY

by

Stefan Brink

Thesis presented in partial fulfilment
of the requirements for the Degree

of

MASTER OF SCIENCE IN ENGINEERING
(EXTRACTIVE METALLURGICAL ENGINEERING)

in the Faculty of Engineering
at Stellenbosch University

Supervisor
Prof. C. Aldrich

Co-Supervisor
Dr. C. Dorfling

March 2015

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 2014/12/05 .....

Abstract

Rock falls are the cause of the majority of mining-related injuries and fatalities in deep tabular South African mines. The standard process of entry examination is performed before working shifts and after blasting to detect structurally loose rocks. This process is performed by a miner using a pinch bar to 'sound' a rock by striking it and making a judgement based on the frequency response of the resultant sound. The Electronic Sounding Device (ESD) developed by the CSIR aims to assist in this process by performing a concurrent prediction of the structural state of the rock based on the acoustic waveform generated in the sounding process. This project aimed to identify, develop and deploy an effective classification model to be used on the ESD to perform this assessment.

The project was undertaken in three main stages: the collection of labelled acoustic samples from working areas; the extraction of descriptive features from the waveforms; and the competitive evaluation of suitable classification models.

Acoustic samples of the sounding process were recorded at the Driefontein mine operation by teams of Gold Fields employees. The samples were recorded in working areas on each of the four reefs that were covered by the shafts of the mine complex. Samples were labelled as 'safe' or 'unsafe' to indicate an expert's judgement of the rock's structural state. A laboratory-controlled environment was also created to provide a platform from which to collect acoustic samples with objective labelling.

Three sets of features were extracted from the acoustic waveforms to form a descriptive feature dataset: four statistical moments of the frequency distribution of the waveform formed; the average energy contained in 16 discrete frequency bands in the data; and 12 Mel Frequency Cepstral Coefficients (MFCCs).

Classification models from four model families were competitively evaluated for best accuracy in predicting structural states. The models evaluated were k-nearest neighbours, self-organising maps, decision trees, random forests, logistic regression, neural networks, and support vector machines with radial basis function and polynomial kernels. The sensitivity of the models, i.e. their ability to avoid predicting a 'safe' status when the rock mass was actually loose, was used as the critical performance measure.

A single-hidden-layer feed-forward neural network with 15 nodes in the hidden layer and a sigmoid activation function was found to best suited for acoustic classification on the ESD. Additional feature selection was performed to identify the optimised form of the model. The final model was successfully implemented on the ESD platform.

Uittreksel

Rotsstortings is die oorsaak van die meerderheid van mynbouverwante ongelukke en ongevalle in diep tabulêre Suid-Afrikaanse myne. Die standaard proses van pretoegang ondersoeke om strukturele los rotse te erken, word uitgevoer voor enige werkskof en na skietwerk. Dit word gedoen deur 'n myner wat 'n breekyster teen die rots kap en 'n oordeel vel op die frekwensie weergawe van die gevolglike klank. Die 'Elektroniese Klinking Toestel' (Electronic Sounding Device, ESD) is ontwikkel deur die WNNR met die doel om die proses te ondersteun. Dit word gedoen deur 'n gelyktydige voorspelling van die strukturele toestand gebaseer op die akoestiese golfvorm gegenereer in die proses van klinking. Die projek stel ten doel om 'n effektiewe klassifikasie-model te identifiseer, te ontwikkel en toe te pas in die ESD om hierdie assessering uit te voer.

Die projek vind in drie stadiums plaas: die insameling van geëtiketteerde akoestiese monsters van die werkareas; die ekstraksie van beskrywende kenmerke van die golfvorms en die mededingende evaluering van geskikte klassifiseringsmodelle.

Klinking akoestiese monsters is opgeneem by Driefontein mynbouoperasie deur spanne van Gold Fields se werknemers. Die akoestiese monsters is opgeneem in werkareas van elk van die vier goudriwwe wat deur die skagte van die mynkompleks gedek word. Monsters is as 'veilig' of 'onveilig' geëtiketteer as aanduiding van die ekspert se oordeel van die rots se strukturele toestand. 'n Laboratorium gekontroleerde omgewing is ook geskep om 'n platform te skep vanwaar akoestiese monsters met objektiewe etikettering waargeneem word.

Drie stelle van kenmerke is onttrek van die akoestiese golfvorms om 'n beskrywende datastel van kenmerke te vorm: vier statistiese momente van die frekwensie verspreiding van die gevormde golfvorm; gemiddelde energie ingesluit in sestien diskrete frekwensiebande in die data; en twaalf 'Mel Frequency Cepstrum Coefficients' (MFCCs).

Klassifikasie modelle van die vier modelsamestellings was kompeterend geëvalueer vir die beste akkuraatheid in voorspellings van strukturele toestande. Klassifikasie modelle het k-naaste bure, selforganiserende kaarte, besluitnemingsbome, lukrake woude, logistieke regressie, neurale netwerke en steun-vektor masjiene met radiale basisfunksie en polinominale kerne. Die meting van die sensitiwiteit van die modelle, met betrekking tot die vermoë van die modelle om veilige voorspellings te beperk wanneer die rotsmassa los is, was gebruik as 'n kritiese werksverrigtingsmeting.

'n Enkel-verskuilde-laag neurale netwerk met 15 nodes in die verskuilde laag en 'n sigmoïde aktiveringsfunksie is gevind as die mees geskikte vir die ESD. Addisionele keuse van kenmerke is uitgevoer deur die geoptimiseerde vorm van die model te identifiseer. Die model was suksesvol geïmplementeer op die ESD platform.

Acknowledgements

I would like to express my sincere gratitude to the following people and organisations.

Dr Dorfling for his assistance and support in realising this dissertation.

Prof Aldrich for his guidance in the choice of the research topic and assistance in the technical aspects of the project.

Gold Fields for their generous assistance in collecting the primary data for this research, and Ras Botma in particular for his role as liaison and team lead.

Beverlie Davies for the proofreading of this dissertation. Any remaining mistakes are purely mine.

Craig Demosthenous for his love and support throughout the project.

My mother for her continued encouragement and faith in me.

My father for his technical advice in all things mining-related, as well as being the main agent in me starting and accomplishing this project.

Toewyding

Hierdie tesis word opgedra met trots aan my pa

Contents

List of Figures

List of Tables

Nomenclature

Acronyms

AE	Acoustic Emission
AUC	Area Under Curve
CART	Classification and Regression Tree
COM	Computer on Module
CSIR	Council for Scientific and Industrial Research
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DWT	Discrete Wavelet Transform
ERB	Equivalent Rectangular Bandwidth
ESD	Electronic Sounding Device
FN	False Negative
FP	False Positive
kNN	k-Nearest Neighbour
LAD	Linear Discriminant Analysis
LTA	Long-Term Average
MFCC	Mel Frequency Cepstral Coefficient
PCA	Principal Component Analysis
RBF	Radial Basis Function
RFE	Recursive Feature Elimination
ROC	Receiver Operator Characteristic Curve
SOM	Self-Organising Map
STA	Short-Term Average
SVM	Support Vector Machine
TN	True Negative
TP	True Positive
VCR	Ventersdorp Contact Reef

Symbols

σ	Standard deviation
θ	Weight coefficient used in logistic regression
C, C_p, λ	Complexity parameters
$K(x, x')$	RBF kernel for SVM
$K(x, y)$	Polynomial kernel for SVM
$M(f)$	Mel value as a function of frequency
r_{xy}	Pearson's Correlation Coefficient
S_a	Spectral distribution skewness
S_c	Spectral distribution mean
S_f	Spectral distribution kurtosis
S_w	Spectral distribution width
$S(z)$	Sigmoid function of given input z
u_i	Spectral moment of order i
k	Generic iterative counter
j	Self-defined iterative counter for Bins
n	Generic maximum indicator
$A(k)$	Amplitude of the k^{th} component of the Fourier transform
$P(k)$	Periodogram value of frequency bin k
$S(k)$	Complex sinusoidial component k
$B(j)$	Average energy value across frequency range of Bin j
$H_m(k)$	Triangular filter function for MFCC
DCT_k	Discrete cosine transform value for component k
$X_{feature}$	Input feature vector
X_{norm}	Normalised input feature vector
X_{stand}	Standardised input feature vector

Chapter 1

Introduction

Rock fall events are responsible for the majority of mine deaths in South African mines. Relatively small rocks that detach from the hanging wall during periods when people are working in the vicinity pose a lethal risk, and contribute to underground fatalities in the South African mining industry.

This continual workplace risk to miners is addressed by various standard mining practices, one of which is the entry-examination procedure. Entry-examination consists of a team of miners evaluating the underground working area after disruptive events such as blasting, and attempting to make the area safe before the rest of the mining personnel are allowed to access the site. Members of the entry-examination team evaluate the roof of the working area, known as the hanging wall in mining parlance, with the use of a pinch bar. A pinch bar is similar to a standard crowbar, and is used to both 'sound' the rock and to bar it down safely if it is found to be dangerously loose. 'Sounding' a rock is the evaluation of the rock's structural stability by judging the noise generated when the rock is struck by the pinch bar. A 'hollow' sound would indicate that the rock might be loose, whereas a 'solid' sound could indicate that the rock is firmly attached to the hanging wall. This process, known informally as the sounding process, has been a mainstay of entry-examination in deep hard-rock mining for many decades. This research will attempt to develop a reliable acoustic classification system to assist the mining personnel in this process to alleviate the chance of errors during rock evaluation.

This chapter details the necessary background to the problem, as well as the research questions and objective addressed in this dissertation. This introductory chapter also provides an overview of the structure of the dissertation.

1.1 Problem background

The Department of Minerals and Resources of South Africa has mandated the capturing of statistics of all mining-related fatalities (MHS, 1996). Examination of the statistics reveals that 39% of mining-related fatalities occurred due

to rock-fall and rock-burst events. Figure 1.1 shows the aggregated accident statistics from 2005 to 2010 and provides a breakdown by accident type.

Managing rock-fall risk is an important aspect of the the deep-rock mining industry. This is a particular concern in South African mines, where the great depth of mining can result in significant pressure on underground excavations (Naidoo, 2007). The main means of excavation is blasting of the ore-bearing rock on a near-daily basis, resulting in a continual redistribution of rock stresses in the typical mining stope. Techniques to prevent rock falls include the deployment of supporting columns and of netting to catch detached rocks, and thorough entry-examination procedures (Poggenpoel, 2012).

The process of entry examination occurs at the start of each mining working shift. Designated responsible miners are tasked to examine the area where the rest of the team will be deployed for hazards. Importantly, all areas between the working face and the first line of support in current workplaces where mining takes place must be examined before a working shift because this is where human activity will be concentrated. Examination includes the identification of the presence of harmful gases, damaged support structures and unstable rocks on the hanging wall and sidewall. The identification of rocks that seem unstable is initially done through visual inspection and then followed by the process of acoustically evaluating the rock's structural cohesion with a pinch bar.

Experienced miners familiar with the sounding process know that an intact

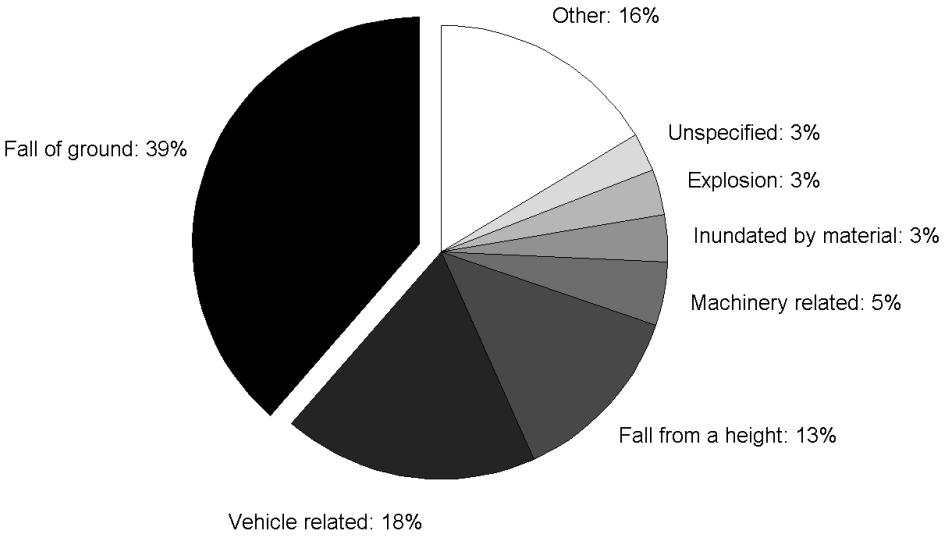

Figure 1.1: A breakdown of causes of mining deaths between May 2005 and March 2010

rock mass, i.e. a rock mass that is considered sufficiently stable to be regarded as safe, will respond to the applied tapping with a relatively high-frequency sound. A detached rock mass, i.e. a rock mass that is insufficiently stable and therefore regarded as unsafe, will respond to the applied tapping with a relatively low-frequency sound (Allison and Lama, 1979). Accordingly, if an experienced miner hears a low-frequency response, he will categorise the rock mass as hazardous, and will usually stipulate that the rock mass must be barred down to detach loose fragments before the area can be declared safe.

The Council for Scientific and Industrial Research (CSIR) has designed and implemented an Electronic Sounding Device (ESD) to provide an additional concurrent assessment of the rock in the sounding process. The ESD records the acoustic waveform generated from the rock and pinch bar during the sounding process. It is designed to classify the structural state of the rock as either 'safe' or 'unsafe', based on the prediction of a classification model listening to the tapping of the sounding bar. After classification, feedback is audibly given to the miner to assist in his judgement of the rock's structural safety.

One of the main design goals of the ESD is for the device to be portable and compact in order to be accepted by the miners. It is envisioned that it should be possible to use the device to determine the integrity of the rock mass without any additional effort on the part of the miner. The ESD is adapted to be mounted on a miner's hard hat, as shown in Figure 1.2.

The sounding process is currently dependent on the miner's attention and experience, and it is hoped that the inclusion of the ESD would mitigate human mistakes during the process. The sounding process forms an essential and important part of entry-examination, therefore the ESD was developed with the goal to improve the detection rate of loose rocks and thereby prevent fall of ground accidents.

1.2 Problem statement

A suitable and effective classification model to perform predictions of the rock structural state based on the acoustic waveforms collected from the sounding process has not yet been developed for the CSIR's ESD.

Prior to this research, the ESD as a hardware platform has been designed and the first functional iteration of the hardware has been prototyped. However, the functionality of the ESD in its stated goal of assisting the miner with an additional structural state prediction during the sounding process has not yet been realised, due to the undeveloped classification system.

Figure 1.2: Photo showing the how the ESD is attached to a miner's hard hat for operational usage

1.3 Research objective

In the light of the above problem statement, the research objective of this dissertation is as follows:

The development of an effective classification system that is able to determine the stability of rock hanging walls in underground mining environments during entry examination based on the acoustic signal generated in the sounding process

The realisation of the research objective is the measure of whether this research is considered successful. The intent is for the classification model to be successfully deployed on the ESD, in order to make the device an effective tool for use in rock-fall risk management in the mining industry.

1.4 Research goals

A number of specific goals were set for this research project. These are listed below. The purpose of this research was to successfully address these goals, in order to meet the overall research objective.

Gather a dataset of labelled acoustic data from the sounding process in a mine-wide geotechnical area: The classification models should be trained and tested against a collected labelled dataset of *in situ* acoustic recordings from the sounding process. An acoustic sample is considered labelled if it is accompanied by an expert's judgement on whether the sound is indicative of a loose or solid rock in a rock mass. Each acoustic sample must be gathered by the current generation of ESD hardware.

Derive a descriptive representation of the acoustic data for the use in classification models: Acoustic waveforms are not suitable for use in classification models directly. The data need to be transformed into a reduced representation for this purpose. A representative feature set needs to be defined that describes the pertinent information and structure in the data.

Find an effective classification model for the judgement of rock integrity: The main goal of this study is to research and implement various possible analytical techniques or models that could enable the ESD to fulfil its intended function. Various concerns were taken into account in choosing and benchmarking these classification models. The benchmark for a successful model is the percentage correlation between the prediction generated by the technique and the judgement of an experienced miner when hearing the same acoustic signal. Correlation should be above 90% before the ESD can be considered viable as a industry-used tool.

Implement the classification model successfully on the ESD: The classification model developed by this project needs to be implementable on the ESD. This means that the technique should not be too computationally expensive for functioning on the ESD platform. In addition, the ESD is only capable of executing applications written in specific programming languages, and the chosen model needs to be implemented in one of these.

1.5 Delineation and limitations

This section lists the necessary delineations to the research objective, as well as the limitations identified in the realization of it.

This research does not aim to develop a new implementation of a classification model. The domain of classification models is currently a very active research area, with a wide variety of well-performing classification implementations in both the public and proprietary spheres. This research will evaluate existing classification models specifically on the acoustic data collected by the ESD and choose the best-performing classifier.

The stability of rocks in the rock wall cannot be objectively measured in operational conditions, except by destructive evaluation. Often the prediction

of structural stability can be verified by barring down the rock under consideration, but this is not always an option. Mining teams often only visually mark a seemingly structurally loose rock for avoidance or further evaluation. Acoustic samples collected in this research were labelled by the prediction from an expert in the domain, but this cannot be considered a true objective measure of structural stability. When the performance of the chosen classification model is presented, its accuracy will be a measure of the correlation between the expert's judgement and that of the classification model.

The sole recording instrument for this research is the ESD. This limits the representation of the acoustic waveform generated by a sounding event to that recordable by the microphone on the ESD.

The limitation on the methodology of this research is the collection of labelled acoustic samples. As will be explained in Chapter 5, the collection of the acoustic samples for this research was done in collaboration with the Gold Fields mining group. The data were collected in a limited period with a modified version of the ESD capable only of recording and labelling samples. Further collaboration with Gold Fields in the testing of the final classification model was not financially feasible for the purposes of this research. To mitigate this limitation, the collected acoustic dataset was partitioned into a classification model training set, and a hold-out independent testing set that was used to evaluate the final performance of the chosen model.

The ESD as it currently exists is the agent for implementation of the classification system. It will still need be developed (reiterated) into a finished commercial product. The classification model could also eventually be deployed to other CSIR projects. However, the processing power on the current iteration of the ESD design is considered a limitation on this research.

Rock falls in the mining industry are not just due to loose rocks that can be detected in the sounding process. The ESD is designed to be used only during entry examination when the sounding bar is used for the detection of loose rocks. The classification model developed in this research will only be effective in this regard.

1.6 Assumption

The assumption of this research is that the sounding process is consistent. That is, a pinch bar of consistent material and length is used only in the evaluation of the hanging wall for loose rocks. This corresponds to the stated aim of the sounding process, but could differ in practical mining situations where individuals use the sounding bar as an all-purpose testing tool during entry examination.

The importance of the strike angle, the location of the strike and the impact strength of the pinch bar striking the rock is key to deliver a consistent data set.

It is assumed that during the collection of the primary data for this research that these aspects are kept uniform throughout the collection process.

1.7 Significance

Realisation of the research objective of this dissertation will enable the ESD to be developed into a vital addition to the process of sounding during entry examination. The use of an effective ESD will ideally mitigate a large component of the inherent risk of human error in the sounding process. The more thorough and objective examination of loose rocks in a mining environment enabled by the use of the ESD will mean that there will be less chance of dangerous loose rocks being missed in the examination. Loose rocks regularly cause serious or even fatal accidents in South African mines, and the ultimate goal is to lessen the prevalence of these events.

1.8 Mining terminology

In this dissertation mining-related terms will be used that are not considered standard knowledge for the reader outside the domain of mining. This section gives an overview of the typical South African hard-rock gold and platinum mines. It then focuses on the process of entry examination used in this type of environment and the terminology related to this process.

A mine, as the term is used in this dissertation, is a shaft that is sunk vertically from the surface in order to reach the ore reefs. From this shaft, horizontal haulage tunnels are excavated at various levels. The levels are between 60 m and 200 m apart and on each level cross-cuts are excavated horizontally from the haulage tunnels to intersect the reef. Figure 1.3 shows the cross-cuts meeting the reef and the raises that are tunnelled to follow the angle of the reef. Figure 1.4 shows how mining proceeds from two raises inwards along the strike in stoping panels. Mining usually stops just before the panels meet in the middle, leaving a dip pillar to support the hanging wall (Green *et al.*, 2010).

Mining in hard-rock mines is predominately cyclical, consisting of a drill-blast-clean cycle. The cycle starts with the drilling of blast holes in the stope face. After all the blast holes have been drilled and filled with explosives, all the workers exit the mine and the explosives are detonated. Once the explosives' fumes have dissipated and seismicity has decreased back to background levels, the mine workers re-enter and begin the process of entry inspection and support installation.

According to Poggenpoel (2012), entry inspection involves checking the hanging wall of the stope for loose rocks and either supporting or barring down those rocks that are found to be unstable. After entry inspection the

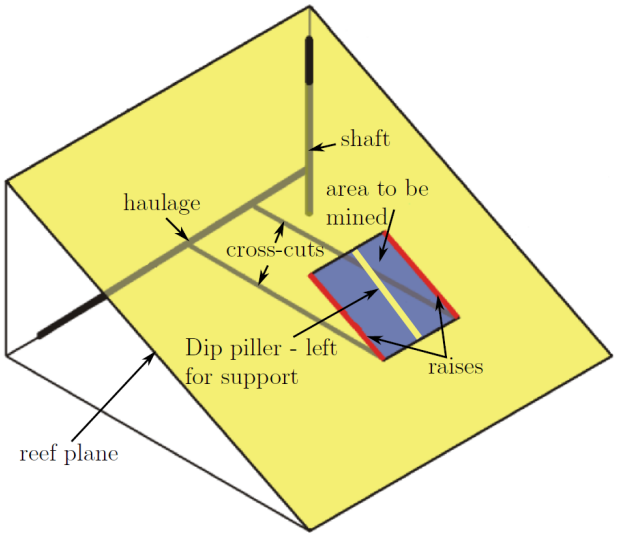

Figure 1.3: Shaft and stope layout of typical deep South African hard-rock mines

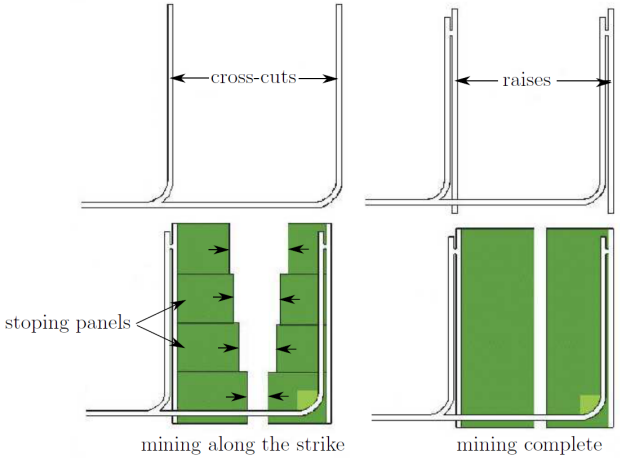

Figure 1.4: The life-cycle of stope mine progression

broken ore from the last blast is removed from the stope using scrapers and then loaded into haulage equipment. In gold mines the haulage equipment is usually rail-bound and the haulage locomotive transports the ore to a tip point where it falls to the bottom level of the mine and is then hoisted to the surface for processing.

The mining-related terms used in this dissertation are listed below, with a short definition of each (Hartman and Mutmansky, 2002).

Ore: a mineral deposit that has sufficient value to be mined profitably

Stope: the working area in an underground mine from which the ore is extracted

Reef: a body of rock containing the ore

Raise: excavations that follow the angle of the reef, from which mining of the reef proceeds
Hanging wall: the 'roof' of a tunnel or stope
Sidewall: the standard walls of the stope
Reef face: the sidewall of the stope which is excavated to obtain the ore
Footwall: the 'floor' of a tunnel or stope
Pillar: a section of rock left to support the hanging wall of a mined-out void
Support: a supporting column made from wood or mining waste between the footwall and hanging wall
Bar down: to prise down a loose rock from the hangingwall with a pinch bar.

1.9 Chapter overviews

Chapter 1 introduces the research project, states the research problem, the goals and objective, the limitations and assumption, and the project's significance.

Chapter 2 is the Literature survey in which current work in the domains of mining, classification models and acoustic features are reviewed.

Chapter 3 details the Methodology for this research, determining the order of the following chapters. The ESD is the main research instrument for this work, and is therefore discussed in detail in Chapter 4.

Chapter 5 provides the specifications of the acoustic data used in this research and the controlled environment from which the experimental acoustic data were collected. The collection process and the mining sites where the operational acoustic data for this research were collected are also described.

Chapter 6, titled Feature data derived, details the transformation of the acoustic data into the representative feature data necessary for input into the classification models.

Chapter 7 competitively evaluates the classification models to determine the most suitable model for use on the ESD.

The Project Conclusion chapter, Chapter 8, summarises the results obtained in the previous chapters and compares these results with the research goals and objective for this research.

Chapter 2

Approaches for acoustic structural analysis

2.1 Introduction

This chapter reviews the domains of mining, acoustic transformation and classification models as applicable to acoustic structural analysis. The aim of this chapter is to evaluate the state of the art in these domains, and provide an overview of the techniques that were used in this research.

The literature review is structured according to the dissertation layout. Section 2.2 reviews mining practices regarding the sounding process. This provides context for the initial phase of the project which focused on the collection of acoustic samples from underground environments that are obtained during the process of sounding.

The second part of the research is the evaluation and transformations of the acoustic samples collected. Section 2.3 surveys the literature on acoustic analysis and manipulation for use in classification models.

Lastly, the research is focused on evaluating classification models for use on the ESD. Section 2.4 provides the theoretical background to and current use of the classification models assessed in this research.

2.2 Mining context

2.2.1 Hanging wall rock mechanics

The rock fall risk risks in deep rock mining are related to the behaviour of the rock surrounding excavations under high confining pressures. This is a particular concern in South African gold mines, where tunnels and stopes are excavated at depths of up to 4 000 m. The failure of hanging walls has been an area of extensive research in rock mechanics as it is the cause of many rockfall-related injuries and deaths (Malan, 1999).

Rock is an inherently variable material in regards to its homogeneity. It can contain many different forms of discontinuity. Small scale pores and crystal structures gives rise to micro discontinuities, whereas large scale faults are due to the boundaries of the grains which introduce features such as joints, extension veins and faults in the rock mass (Bosman *et al.*, 2000). In South African deep tabular mines where blasting is the main form of excavation, these discontinuities manifest as fractured zones around the excavation. The closure phase of the excavated areas in the period after blasting exaggerates stresses on existing discontinuities, which can lead to rock bursts and fall of ground events of different magnitudes in the working area.

A non-destructive evaluation method of discontinuities in a rock mass is by means of acoustic emission (AE) (Lockner, 1993). The propagation of acoustic waves through rock is characterised by an elastic displacement of the material. This is directly influenced by the degree of elastic coupling between the different media through which the acoustic wave propagates. A rock with discontinuities presents a mixed-media environment for a propagating acoustic wave, resulting in a higher reflection rate of the acoustic wave. This is characterised by a higher portion of the impact energy spectrum being reflected to the source of the impact (Tarantola, 1984).

Lower frequencies propagate more readily in an elastic medium such as rock. A rock mass without discontinuities will attenuate higher frequencies to a greater extent than a heterogeneous rock mass. This is known as anelastic attenuation in reflection seismology (Gladwin and Stacey, 1974). Over significant distance rock as a propagation medium can be seen as a low-pass filter. The size of the homogeneous environment is therefore inversely proportional to the presence of lower frequency in the reflecting wave. This can be empirically observed in the case of entry-examination, where the sounding response from a loose rock will typically contain more non-absorbed lower frequency energy.

The presence of discontinuities such as fractures in hanging walls are directly related to the various rock types present. The different ore-bearing reefs that are excavated in the South African mining context are associated with deformation after blasting. Denkhaus *et al.* (1958) found that the susceptibility to rock failure was more prevalent in stopes in the Central Rand mining region where the reef lies between hard brittle quartzite bands. This condition can be seen in two of the reefs that are assessed in this dissertation; the Ventersdorp Contact reef (VCR) and the Carbon Leader reef. According to Schweitzer and Johnson (1997) the hanging wall of stopes on the VCR can be hard and soft lava, the differing strengths of which varies greatly from area to area. The presence of the 'soft' lava results in highly fragmented hanging wall conditions. Hanging walls on the Carbon Leader reef can consist of siliceous quartzite, the Rice Pebble Marker or the shaly Green Bar. From this description it is clear that a univariate characteristic of the rock condition in hanging walls should not be assumed even in a single reef type.

The hanging wall in operational stopes experience large stresses ahead

of the working faces where blasting takes place. These normally exceed the strength of the rock, resulting in the formation of an extensive fracture zone. Under high stress conditions, this zone can extend to more than 5 meters ahead of the face. Adams *et al.* (1980) found that the majority of rock fracturing in the hanging wall occurs in the relatively short period after blasting. In is in this area where human activity is most concentrated during the cleaning and preparation phases of the cyclical mining cycle and therefore the process of effective acoustic entry examination is important to establish the safety of the hanging wall.

2.2.2 Entry examination technologies

The part of the entry examination that has a direct bearing on this research is the process of sounding. Numerous studies have been done on the analysis of the frequency response of rocks in various states of structural cohesion (Summerfield, 1956). Allison and Lama (1979) examined the methods in use for detecting loose rock bars, and developed a theoretical framework for frequency response changes in rock separation cases. Allison and Lama point out that early detection and remedial action could save time, money and, in many cases, lives. Allison and Lama also refer to the fact that many different techniques to test for loose rocks have been attempted, but that the most common technique remains listening to the rock struck by a hammer. Allison and Lama experimented with a technique that can determine the rock parameters by applying ultra-low-frequency vibrations of a small magnitude to the rock, and detecting the response with a sensor attached to the rock. Classification results were not provided in this study apart from a frequency response figure. A shortcoming of this technique is the preparation of the rock to attach actuators and sensors to it, a process that would have to be repeated for each rock tested.

Hanson (1985) came to the conclusion that a system developed which adds additional preparation work during entry examination is likely not to be adopted by the mining community. Hanson did a study on the convenience, simplicity and accuracy of techniques that can be used to provide an objective measure of rock stability. The techniques assessed were low-frequency and ultrasonic acoustic sounding, radar probing and high-resolution resistivity. Hanson deemed low-frequency acoustic sounding to be the most promising of these techniques, and proceeded to develop instrumentation to use this analysis technique. The device developed was patented by Hanson (1986). The instrumentation consists of an accelerometer mounted on the pinch bar connected to two band-pass filters which calculates the energy in a defined high and low band of spectra. The frequency ranges chosen were [500 Hz–1 000 Hz] for the low-pass band and [3 000 Hz–3 500 Hz] for the high-pass band. The comparison of the two aggregate energy values is then used to determine the state of the rock. This research and the patent have close similarities to the

objective of this dissertation. Hanson (1986) does not report quantitative results, however successful classification by the device is claimed. The main shortcoming of Hanson's device is the strictly defined frequency ranges evaluated. The definition of the spectrum bands was derived from coal mining environments, which differ significantly in rock type and depth from gold and platinum hard-rock mines in South Africa. Furthermore, while the patent does circumvent the requirement to pre-prepare the rock surface, it still relies on mounting an accelerometer on the pinch bar itself. This demands a specially adapted pinch bar, which would have to be present and maintained at each site during entry examination.

Lux *et al.* (1991) patented a loose rock detector with the change in innovation being the implementation of variable defined frequency ranges to compare with each other. The patent does not state the success rate in classification by the device. The prerequistite of explicitly mounting the accelerometer on the rock is also present in this patent. The benefit of Lux *et al.*'s approach over that of Hanson is the inclusion of memory on the processing platform, which enabled post-event analysis.

Carceles and Muller (2007) detail the use of classification models for categorising rock events. This is an approach that is followed by this dissertation. The scope of classification in the work by Carceles and Muller is focused on a broad range of seismic events, and not on the identification of loose rocks in particular. It does show the applicability of classifying time-based waveforms related to mining, in this case seismic waveforms, into discrete classes. The work shows that the use of features extracted from the time-based waveforms is essential as a prerequisite for successful classification. The system described by Carceles and Muller classified French regional seismic events with an overall performance of about 90%, with a specific accuracies of 86% for earthquakes and 91% for explosions and approximately 99% accuracy in the detection of rock-bursts.

The work by Schultz *et al.* (2008) is an additional example of the use of comparison in the frequency domain to detect the failure rate in underground environments. The failure being monitored is in the state of a drill operating on rock in real time. The comparable aspect to this research is in the successful demonstration of failure detection from waveforms generated by a tool operating on rock. The pinch bar used in the sounding process is potentially simpler to evaluate than a drill in operation.

2.3 Acoustic features

2.3.1 Feature importance

In the previous section, technologies deployed in the assessment of rocks were based on descriptive acoustic features that are exclusively derived from the

frequency domain. The features used were derived from methods akin to band-pass filters, resulting in total or averaged spectral energy values obtained from specific frequency band ranges. Spectral energy values constitute only a part of available acoustic features that can be derived from audio waveforms. However, the bulk of the literature regarding acoustic features is found in other research domains. The task is then to find acoustic features from these other domains that can be used in this project.

Audio data is a prevalent part of many computer and integrated device applications. In particular, audio data used in classification models forms an integral part of voice recognition, music genre classification, voice-command systems and various monitoring applications. Due to the extent of audio data used as the main input to these systems, a large library of literature exists on the use of acoustic data in classification systems. This section reviews the pertinent literature from these domains applicable to this research.

A note on the terminology used in this section. A feature of a sample is a single-valued descriptor of the sample. Further on in this dissertation, this concept is expanded on, particularly in Chapter 6 which derives the features used in this dissertation. Features are used as an intermediate transformation between raw acoustic samples and the classification models that make predictions based on the samples.

The most comprehensive taxonomy of available acoustic features were found in the work by Moerchen *et al.* (2004), which was presented by Moerchen *et al.* (2006). The research performed by Moerchen *et al.* were done in response to a perceived shortcoming in the literature in regards to an exclusive focus on acoustic features. As Moerchen *et al.* state, research in automatic speech recognition and music classification emphasises the choice and implementation of classification models, rather than the acoustic features. They proceeded to do an investigation exclusively into audio features. The researchers firstly provided a useful taxonomy of existing audio features, separating it into two categories: low-level features and high-level features. Low-level features are those that are derived from short time windows, typically in the order of 20 ms to 200 ms. High level features are products of low level features over longer periods, providing a description of how the signal changes over time.

The ESD performs its prediction on short time window acoustic phenomena due to a recorded sample being 180 ms in duration. The relevant part of the research by Moerchen *et al.* to this project is therefore the low level features. Nine groupings of low level features are presented in the research paper, derived from the authors' review of the state of the art in domain of acoustic features. These are shown in Table 2.1 on the following page.

The nine groupings are as follows:

Elementary descriptors Features that are calculated in the time-domain, as opposed to the other features which are derived from the frequency domain. These features are primarily related to the volume of the signal.

Table 2.1: Low level acoustic features as derived by Moerchen *et al.* (2004)

Group	Feature
Elementary	Volume
	Zerocrossing
	Lowenergy
Spectral features	Spectral Centroid
	Spectral Bandwidth
	Spectral Rolloff
	Band Energy Ratio
	Spectral Crest Factor
	Spectral Flatness Measure
	Spectral Flux
Linear regression	SpecReg Slope
	SpecReg Y Intercept
	SpecReg Maximum Error
	SpecReg Medium Error
Peak determination	SpecPeak Amplitudes
	SpecPeak Frequencies
	SpecPeak Widths
Pitch content	Pitch Content
Psychoacoustic scales	Mel Magnitudes
	Bark Magnitudes
	ERB Magnitudes
	Octave Magnitudes
Cepstral coefficients	MFCC
	BFCC
	EFCC
	OFCC
Chroma	Chroma
	Normalized Chroma
	Mean Chroma Tone and Strength
Advanced Bark	Bark/Sone Loudness

A recurrent feature elimination criterion is first encountered here: these features are not relevant to this research since the volume of a waveform generated by a rock struck with a pinch bar is assumed to be highly variable and therefore non-informative due to noisy mining environments.

Spectral features Features that describe the shape of the distribution. These features are used in this project, and are described in Subsection 2.3.2. However, the spectral flatness and crest factors features were not included in this project due to the expectation that sinusoidal waveforms were not expected from impacting rocks. Another omission from this research is the spectral flux feature , which is a measure between subsequent short waveforms and is not applicable to this research in which only individual waveforms are analysed.

Linear regression Features that describe the shape of the frequency distribution as approximated by a linear regression model. The authors do not describe the method used to derive these values apart from a reference to a paper which is similarly opaque on the topic (Mierswa and Morik, 2005). It has been decided not to pursue this grouping as a candidate for this project.

Peak determination Features which indicate where the significant peaks are in the frequency distribution. These are based on an algorithm devised by the authors that determines the location of the 5 largest peaks in the signal. The same rejection criterion from the previous grouping applies here; the implementation details of the method is not provided in the original paper nor the referenced papers.

Pitch content These features use the positions and amplitudes of the three most prominent peaks of the enhanced autocorrelation represented by corresponding stored musical tones (in MIDI files). This can not be replicated in this research due to the fact that no similar templates of tones exists for sounding events.

Psychoacoustic scales Features representing the spectral energy in ranges informed by human hearing perception. These scales are used in this research. A description and justification of this this grouping is given in Subsections 2.3.3 through 2.3.6.

Cepstral coefficients Features that represent an additional transformation of the psychoacoustic scales. These features are used in this project. An introduction to the usage is provided in Subsection 2.3.7.

Chroma These features are a variation on the octave scale which forms part of the psychoacoustic scales discussion. As will be discussed in Subsection 2.3.4, the octave scale is not chosen for this research due to its relative lack of resolution in the frequencies useful in distinguishing loose rocks.

Advanced Bark The Bark/Sone loudness scale is an advancement of the Bark scale. The Bark scale is not used in this research due to its derivation based on volume perception. As was previously mentioned, volume

is not a measure that is considered informative in the prediction of loose rocks in noisy environments.

The features used in this research are therefore a subset of the low level acoustic features evaluated by Moerchen *et al.* (2004). They discussed the usage of these features in the problem domain of music genre classification, which apart from a similarity to audio-based classification does not share further similarities to this research. The work presented does however provide a comprehensive roadmap to the choice of potential features that can be adapted to this project.

Gillet and Richard (2004) devised a classification system for the automatic transcription of drum loops. This system was based on the analysis of percussive acoustic events, which bears similarity to the impact acoustic events assessed by the ESD. The authors succeeded in implementing the system based on a grouping of low level acoustic features. These were a combination of spectral features, cepstral coefficients and a modified version of the Mel psychoacoustic scale. This same approach of choosing useful existing features was adapted in this research as an alternative to pure frequency band content features as was used in the loose rock detection technologies discussed in the previous section.

A full discussion of the acoustic features used in this project is provided in the following subsections.

2.3.2 Spectral features

Features that describe or aggregate the frequency distribution are known as 'spectral features'. Descriptors of a frequency distribution that can be used as features are the statistical measures of mean, deviance, skew and kurtosis. In terms of acoustic features, these features are alternatively known as the spectral centroïd, spectral width, spectral asymmetry and the spectral kurtosis (Grey and Gordon, 1978). These are directly related to the first four order statistical moments, and are given as follows:

Spectral centroïd (or mean)

$$S_c = \mu_1 \tag{2.3.1}$$

Spectral width (or variance)

$$S_w = \sqrt{\mu_2 - \mu_1^2} \tag{2.3.2}$$

Spectral asymmetry (or skewness)

$$S_a = \frac{2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3}{(S_w)^3} \tag{2.3.3}$$

Spectral flatness (or kurtosis)

$$S_f = \frac{-3\mu_1^4 + 6\mu_1\mu_2 - 4\mu_1\mu_3 + \mu_4}{(S_w)^4} + 3 \tag{2.3.4}$$

where the i^{th} spectral moment of order is given by $\mu_i = \frac{\sum_{k=0}^{N-1} k^i A(k)}{\sum_{k=0}^{N-1} A(k)}$ in which $A(k)$ is the amplitude of the k^{th} component of the periodogram estimate of the acoustic waveform.

A periodogram is a standard signal transform that is used to represent an estimation of the spectral density in a time-series signal. This is accomplished by applying a Discrete Fourier Transform (DFT) to the signal. For a signal, $s(n)$, with n data points, the result of the complex DFT is called $S(k)$. The DFT is applied by the following equation over each sample where K is the length of the DFT:

$$S(k) = \sum_{n-1}^{N} (s(n)e^{-j2\pi kn/N}) \quad 1 \leq k \leq K \tag{2.3.5}$$

Each resulting $S(k)$ value of a DFT results in a complex number that represents both the phase and the amplitude of a sinusoidal component of $s(n)$. If $s(n)$ was a complex function, $S(k)$ would result in an unsymmetrical distribution. An interesting characteristic of Fourier transforms is that when applied to real signals, such as waveforms collected by the ESD, it results in an $S(k)$ distribution that is completely symmetrical (Sorensen *et al.*, 1987). This relationship is captured as follows:

$$S(k) = S^*(N - k) \tag{2.3.6}$$

This leads to a useful corollary of the mentioned characteristic — due to both halves of the $S(k)$ distribution being identical, only half of the resulting $S(k)$ needs to be considered for practical frequency analysis.

After $S(k)$ has been derived and halved, the absolute value of the complex Fourier transform is taken and the result is squared to obtain the periodogram estimate of the acoustic signal. This final transformation is given in the following equation:

$$P(k) = \frac{1}{N}|S(k^*)|^2 \tag{2.3.7}$$

2.3.3 Psychoacoustics

The domain of psychoacoustics provides a justification for the use of defined frequency scales that can be used to aggregate ranges of frequencies into discrete average energy values called bins. These aggregated bin values can then be used as features for classification models. The human hearing system can

detect frequencies in the range of 20 Hz to 20 000 Hz, but due to the physical structure of the ear canal, the human ear is especially sensitive between 2 000 Hz and 5 000 Hz according to the tests done by Takeshima *et al.* (2001). For the purposes of this research, mapping feature options that resolve this critical part of the frequency domain with a high degree of resolution will be prioritised. Due to the close correlation of this frequency band to the human experience of pitch, as defined below, existing pitch related feature mapping options are examined in this section.

The human hearing system can be said to perform a spectrographic analysis of any auditory stimulus according to Pickles (1988). This neural analysis of the spectrum is then used by the human brain to determine the characteristics of the source of the sound, as indicated by Zwicker and Fastl (1999). The auditory portion of the inner ear, the cochlea, can even be regarded as a bank of filters whose outputs are ordered tonally.

This research wants to approximate the human hearing system in the feature mapping step of the system. For clarity of reading, it is necessary to distinguish the concept of pitch as opposed to frequency. Pitch is an auditory process in which a listener performs an internal assessment of the sound heard to a subjective scale, i.e. each pitch is able to be sorted higher or lower than another pitch by the listener. Pitch is closely related to frequency but the two are not the same - an observation that Stevens and Volkmann (1940) clearly stated. Succinctly, frequency is an objective scientific concept, whereas pitch is subjective to the listener. It takes a human brain to map its own assessment of pitch.

Efforts have been made to derive a quantitative scale representation of pitch. Various auditory scales of pitch have been derived empirically through experiments and subsequently codified as equations. These range from the well-known musical scale representation of pitch to other implementations used in speech recognition and noise analysis. The four different frequency scales in wide use are known as the Octave, Mel, Bark and Equivalent Rectangular Bandwidth (ERB) scales. These will each be described individually. These scales are compared with one another and their applicability and potential implementation in the current research are also compared.

2.3.4 Octave scale

The Octave scale, alternatively known as the musical pitch auditory scale, is a broadly recognised form of scale representation. This scale is divided into the well-known octave notation that is used in musical theory. It forms the basis of various musical instrument design choices, such as a piano's keyboard layout pattern repeating on every octave jump in pitch.

The usefulness of the musical pitch notation to this research stems from the fact that the modern musical pitch scale is logarithmically proportional to frequency with only minor deviations. This means that an effective doubling of

frequency occurs between equivalent musical notes in each increase of octave. A clear example of this is shown in Table 2.2, which gives the frequencies of the A note over the first nine octaves of the scientific musical pitch notation. This scale was first described by Young (1939), and has been subsequently implemented as a standard in ISO 16:1975. The A note on this scale is considered the standard concert pitch for tuning musical instruments. The choice of the A note as an octave delimiter for this research is particularly useful due to its simple frequency values over all the octaves.

The use of musical pitch scales as a feature mapping basis is found mostly in the domain of automatic music genre classification due to the clear overlap in the musical domain. The literature describing this problem is extensive. Tzanetakis and Cook (2002) propose using pitch content as a feature set for statistical pattern recognition classifiers. Poria *et al.* (2013) also use musical pitch scales as a candidate for an extended time feature set, though a combination of sets are used in the final implementation of that design and musical pitch is used as a baseline method to compare the efficacy of other methods.

Musical pitch as a feature set is the simplest implementation of the auditory mapping tools that are discussed in this section. The simplicity of this approach makes it an easily implementable candidate for providing a baseline comparison with other feature mapping methods.

Table 2.2: The location of the frequency spectrum of musical note A in the first 9 octaves of the Western Chromatic Scale according to Young (1939)

Octave	0	1	2	3	4	5	6	7	8	9
Freq (Hz)	27.5	55	110	220	440	880	1 760	3 520	7 040	14 080

2.3.5 Mel scale

In the introduction to this subsection it was noted that human hearing is especially sensitive in the range of 2 kHz and 5 kHz. In Table 2.2 it can be seen that the previously described musical pitch scale only has two octaves covering this important range of the hearing spectrum. It is necessary to identify methods that have more resolution in this range to provide a more rigorous input to classification models on the ESD.

The Mel scale was initially named by Stevens *et al.* (1937) in their paper titled, "A scale for the measurement of the psychological magnitude pitch", and is defined as a perceptual scale of pitches judged by test subjects to be equal in distance from one another. The word *Mel* comes from the word melody to indicate the design of the scale as based on pitch comparisons. In their paper, Stevens *et al.* used a variety of experimental methods and approaches to determine curves based on just-noticeable differences of pitch as indicated by

human test subjects. These curves were not initially published with formulae to approximate them, and therefore subsequent approximations of the curves were attempted in papers by Koenig (1949) and Fant (1949).

The most widely used version of a curve-fitting formula was derived by Makhoul and Cosell (1976), and is given by:

$$M(f) = 1125ln(1 + \frac{f}{700}) \tag{2.3.8}$$

In fact, many authors have attempted to fit formulae for a variety of reasons to the derived Mel curves. Umesh *et al.* (1999) concluded that there are many qualitatively different equations, each with a few parameters, that fit the experimentally obtained Mel scale. In the paper in which this conclusion was drawn, the authors presented the values they obtained from a figure in the paper by Stevens and Volkmann (1940). These values are presented in Table 2.3. This table provides the descriptive breakpoints of the Mel scale that will be adapted in this dissertation.

Table 2.3: Table of measurements by Umesh *et al.* (1999) relating frequency to mel values

f =	40	161	200	404	693	867	1000	2022	3000	3393	4109	5526	6500	7743	12000
F_M =	43	257	300	514	771	928	1000	1542	2000	2142	2314	2600	2771	2914	3228

2.3.6 Other scales

The Bark scale was motivated by the observation that the sensed volume is constant within certain special frequency ranges, but different outside of them. These bandwidths are called the critical bandwidths and were published by Zwicker *et al.* (1957). There are many competing formulations attempting to fit the Bark scale. As previously mentioned, the loudness of the signal is not of particular concern to the research in this dissertation, as widely varying volume is expected in the collection of sounding-related samples. The ERB scale is a modern derivative of the Bark scale derived by Moore and Glasberg (1996), and is also a representation of loudness.

2.3.7 Frequency cepstral coefficients

Each of the scales discussed in the previous subsections is derived from the frequency domain. A useful additional transformation to the output of energy values obtained by using these scales is the Discrete Cosine Transform (DCT). Effectively, this is an additional transformation that is similar to the initial one from the time domain to the frequency domain of the acoustic waveform. This spectrum-of-spectrum transformation results in what literature refers to

as the cepstrum (Childers *et al.*, 1977). Note that 'cepstrum' is a respelling of the word spectrum with the first four letters reversed.

Cepstrum indicates the rate of change of the individual spectrum bands. Applying the DCT to obtain the cepstrum offers a decorrelated description of the strongest trends in the spectrum. The cepstral coefficients are obtained when the transform is applied to areas of the frequencies defined by frequency scales. The resulting coefficients are named after the scale that is used. For instance, if the Mel scale is used, the end product is called Mel Frequency Cepstral Coefficients (MFCC). Similarly, for the Bark scale, the BFCC.

This section will not go into depth about how this process is applied, since a step-by-step derivation of the calculation for the MFCC is performed in Chapter 6, Section 6.5 on page 83.

Frequency Cepstral Coefficients are a ubiquitous measure used in acoustic classification. The coefficients are an effective way to gain additional insight into the source of the acoustic signal. Cepstrum as a concept was originally applied to the categorisation of seismic sources, though it has been found to be not as effective as other techniques in terms of source separation (Childers *et al.*, 1977). It has since found wide use in speech recognition, environmental acoustic classification and music genre identification, as respectively seen in work done by Ma *et al.* (2006), Chu *et al.* (2009), and Ludeña-Choez and Gallardo-Antolín (2014).

2.4 Classification models

The study and use of classification models is a topic that exists under many names in the domain literature. Usually though, the use of the respective names is standard across the specific application domain of the model. In formal statistics, it is often called 'predictive modelling'. In the computer science domain it is well known as 'machine learning'.

Predictive modelling differs from the statistical concepts of descriptive and prescriptive modelling. In the case of descriptive modelling, the data are analysed for insight. In prescriptive modelling, the data are used as guide to constructing models. In predictive modelling, however, predictions are made directly from the data without necessarily forming a deeper understanding of the problem phenomena. This approach by predictive modelling is a strength in that it allows for the construction of models for problems that might be too complex for a conceptual model. The corresponding weakness of predictive modelling is that the output may very likely be a 'black-box' solution with no accompanying increase in conceptual understanding of the problem (Janert, 2010).

Classification models can be categorised into three approaches:

Supervised learning: All data samples used for training are allocated a pre-

defined class label. For instance, in this research the class labels are 'safe' or 'unsafe', referring to the perceived state of the rock. The class labels are explicitly provided and do not need to be inferred from the data during the construction of the classification model. Classification models are trained and tested by their ability to assign samples to the correct class.

Unsupervised learning: In this technique samples are grouped together based on similar information in the samples. The similarity in the data are used to predict the class values of new samples. No class labels are explicitly provided.

Recommendation: These are systems that tend to be used in predicting human behaviour. Models created through this technique recommend a suitable item based on past behaviour. This is not particularly relevant to this research.

Classification models based on acoustic data have been shown to be constructed with either supervised or unsupervised learning techniques. Lamel *et al.* (2002) discuss the benefit of using unsupervised techniques in constructing acoustic classification models to reduce the effort cost in obtaining correctly labelled data with which to train. However, they still have to rely on what they refer to as 'lightly' supervised techniques to initiate their model. The same initial reliance on supervised training techniques for acoustic-related problems was found by Novotney *et al.* (2009) and Fraga-Silva *et al.* (2011) during their development of unsupervised models. This indicates that classification on acoustic data is reliant on the supervised learning approach to the construction of classification models.

Classification models used in the field of acoustic-related prediction are predominantly found in the domains of music genre recognition and human language computational understanding. These are not fields that are directly related to the structural assessment of rocks. The fact remains, however, that the bulk of the research in acoustic-related classification is found in these domains and therefore some of the literature reviewed in the following sections will be from these domains.

The classification system developed in this research bears similarities to process monitoring in the field of chemometrics. Fault detection in industrial systems is the first step in the chain of detection, identification, diagnosis and recovery. Therefore emphasis on classifying faults is necessary. The goal of process monitoring is similar to the goal of the classification system developed in this research, that is to develop fault detection algorithms that are maximally sensitive and robust to all possible faults (Chiang *et al.*, 2001).

The classification models discussed in the following subsections are grouped by common approach, or modality. This is to facilitate the discussion about these models where similar theoretical underpinnings are consistent between

models. There are also the groupings that are used in the presentation of the results of the classifiers' performance in Chapter 7.

2.4.1 Instance classification

Instance-based learning classification models predict a problem by comparing of instances of sample data that are deemed important to the model. Methods in this category typically build up a database of training data. When applied to testing data, the classifier uses a similarity measure in order to find a best match and make a prediction. Due to this, instance-based classifiers are also called 'winner-take-all' methods, or alternatively memory-based learning.

The first classifier evaluated in this grouping is the k-Nearest Neighbour (kNN) classifier. kNN is considered to be the least complex classification model due its ease of understanding and implementation. The kNN classifier does not attempt to construct a general internal model, it simply stores instances of the training data. Classification is computed from a majority vote of the k-nearest neighbours of each point, and the prediction is assigned to the class of the most representatives within the nearest neighbours of that point. The distance metrics between neighbours differ by implementation, but are typically Euclidean or Minkowski distances. Class probability estimates for the new sample are calculated as the proportion of the stored training set neighbours in each class. The new sample's predicted class is the output with the highest probability estimate (Altman, 1992).

Maillet *et al.* (2014) used the kNN approach to create two clusters of samples that determine whether acoustic signals derived from ceramic blocks are indicative of structural damage. They expressed satisfaction at the performance of the kNN classification model, and inspection shows strong separation between the two structural classes.

The second classifier evaluated in this category is the supervised Self Organizing Map (SOM). This classifier is also known as a Kohonen map, based on the name of the researcher who developed the idea (Kohonen, 1982). SOM is not traditionally seen as a supervised classification model; its broad usage is in unsupervised clustering of data. However, supervised SOM systems have been found to perform on par with kNN and advanced methods such as neural network and support vector machines (Gil and Johnsson, 2010).

A Self Organizing Map consists of a set of non-interconnected units which are spatially ordered in a grid-like topology. Each unit in the map is equipped with a weight vector of which the number of elements is equal to the number of variables in a training sample. The weights are initialised by random numbers. Individual samples in the training set are presented to all the units in the network. The unit in the map possessing the weight vector most similar to the presented sample by a certain similarity measure is assigned as the winner. The training algorithm will then update the weights of the winning unit and its neighbours on the map to be more slightly more similar to the input sam-

ple. Winning units in this process becomes specialised to the samples which are frequently mapped onto it, and therefore the unit will respond to similar samples in the future.

A SOM with few units in the structure is equivalent to a kNN classification model, as shown by Hannula *et al.* (2003) in their modelling with frequency domain data. However, due to the non-linear weight relation between the input samples and units in the map, emergent behaviour is observed when more units are added to the map. When the SOM is initiated with enough units in its map, it becomes topological in nature. Researchers exploit this characteristic by implementing novel distance metrics on the observable map in order to gain understanding and predict new samples.

Moerchen *et al.* (2006), in work previously mentioned in the acoustic features section of this chapter, used the SOM approach extensively in order to identify the best acoustic features to use for classification. The researchers extended the use of the SOM model to create new decorrelated features that outperformed many of the standard ones they evaluated.

Tibaduiza *et al.* (2013) successfully used SOM to classify structural damage based on acoustic emissions.

2.4.2 Decision-tree classification

Decision trees are a classification technique that can be very interpretable in simple cases, and used quite widely in applications where the researchers want to know how a feature is specifically interpreted to derive a predicted class. In this role, decision trees and the advanced classification models that are based on them can be used for a deeper understanding of the problem being modelled. The intuition to understand a classification tree is quite simple: the method starts by searching for a single value in the features that best splits the data into two groups that are distinct by their class values. A classification tree does this by dividing the input feature space into recursive binary partitions according to a defined purity measure.

Classification and Regression Tree (CART) analysis was first introduced by Breiman *et al.* (1984), and formed the core of all the development in this area that followed. A decision tree as defined by Breiman *et al.* is used in a classification application with class outcome $\kappa = 1, 2, \ldots, C$, where the following in defined for region R_η with N_η observations in node η

$$\hat{p}_{\eta\kappa} = \frac{1}{N} \sum_{x_i \in R_\eta} I\left(y_i = \kappa\right) \tag{2.4.1}$$

where $\hat{p}_{\eta\kappa}$ is the fraction of samples allocated to node η with class outcome κ, and $I(\bullet)$ is an indication function. Observations allocated to node η are classified as the class corresponding to the majority of of the nodes in that node.

Typically, the best split minimises the impurity of the resultant class distributions in the two data subsets. For the two resulting groups, the method is repeated a set amount of times until a certain information rate is achieved, and results in a hierarchical structure that is visually similar to the layout of a tree. The aim of classification trees is to partition the data into smaller homogeneous groups in terms of the class. Selecting optimal splitting variables and split positions ζ for recursive partitioning is done by applying a certain purity criteria. Breiman *et al.* originally used the Gini index as the purity measure in classification problems. The node impurity measure Q_n is defined as

$$Q_n = \sum_{\kappa=1}^{K} \hat{p}_{\eta\kappa} \left(1 - \hat{p}_{\eta\kappa}\right) \tag{2.4.2}$$

A variation on this split, which is more often used in modern implementations of decision trees, is the split according to information theory. The information statistic is given by the binary entropy function

$$H_b(p) = -\left[p \log_2 p + (1-p) \log_2 (1-p)\right] \tag{2.4.3}$$

where H_b is the entropy of a discrete random variable and p is the probability of the first class. Decision tree classification implementations that use this measure are C4.5 and its variants, initially developed by Quinlan (1993). While CART and C4.5 are the most widely used, there are many different proposals for tree-based classifiers. This dissertation uses CART as the example of a single tree classification model for testing, with the aim of delving deeper into this domain if the initial results are competitive with the other classification models considered in this research.

Random Forest is a powerful evolution of the tree-based classification model approach. Random Forest is a name trademarked by Breiman (1999), who first illustrated the concept, although the classification model is almost never referred to as Random Forest™. The seminal paper by Breiman describes Random Forest as a method of building a forest of uncorrelated CART trees, combined with randomised node optimisation and bagging. Random Forest is an example of a ensemble algorithm.

Breiman suggested training K trees on various combinations of input samples and feature variables. This ensemble of trees uses majority voting to determine the global classification output of the Random Forest algorithm. The variation between individual trees is accomplished by manipulating the learning set of each tree through bagging. Bagging is the term for bootstrap aggregating. Bootstrapping consists of taking a different sized sample set of size n from the entire training set as a modified training set for each new tree. Furthermore, random split selection is deployed for each tree by restricting the available split variables at each decision node to m random feature variables. This method attempts to establish diversity between that ensemble members,

which would mean that the errors they make on unseen data are ideally uncorrelated. When the number of trees in the ensemble is sufficiently large, the generalisation error of the model converges to an optimal limit.

Ntalampiras and Roveri (2013) presented a study that adheres closely to the research objective of this dissertation. They tested for imminent rock collapse in the Alps in Northern Italy through *in situ* monitoring, covering a period of more than one year of micro-acoustic signals. They structured their study around the identification of the best-performing classification models for their data. The classification models evaluated were J48 tree (a variation of C4.5), Linear Discriminant Analysis, Multilayer Perception (a specific type of neural networks), Random Forest and Support Vector Machines with radial basis function kernels. All of these classifiers are considered in this dissertation as well, with the exception of Linear Discriminant Analysis (LDA). Ntalampiras and Roveri found that the smallest number of misclassification in the study was achieved by the Random Forest classifier.

2.4.3 Function classification

In this section the broad classification grouping that depends on function-based classification is evaluated. These classifiers have an activation function as their primary classification mechanism, and are trained by estimating the error in the model during training through an optimisation function. The model complexity stemming from this broad definition can vary widely, as arguably the most interpretable and least interpretable classifiers are both covered by it. These are the classification models known respectively as logistic regression and neural networks.

Logistic regression classifiers are the best-known linear classifiers in the literature. Logistic regression attempts to construct a linear separation between training data samples that distinguishes the samples by class. This separation function is characterised by the sigmoid function:

$$S(z) = \frac{1}{1 + e^{-z}} \tag{2.4.4}$$

For logistic regression, z would be replaced by a weighted relationship of $X_{feature}$, as follows:

$$z = \Theta^T X_{feature} \tag{2.4.5}$$

where X is the input features given by an n-dimensional discrete vector $X_{feature} = (x_1, x_2 \dots x_n)$, and the weight coefficients are given by $\Theta = (\theta_0, \theta_1, \theta_2 \dots \theta_n)$. Finding the best values for Θ in order to separate the classes of the problem would be the goal of any model training implementation. The sigmoid function results in an S-curve, shown in Figure 2.1, which is bounded by $[0, 1]$ over the entire possible range of $X_{feature}$. This property is helpful in limiting the output

of logistic regression to the same probability distribution used for evaluating other classification models.

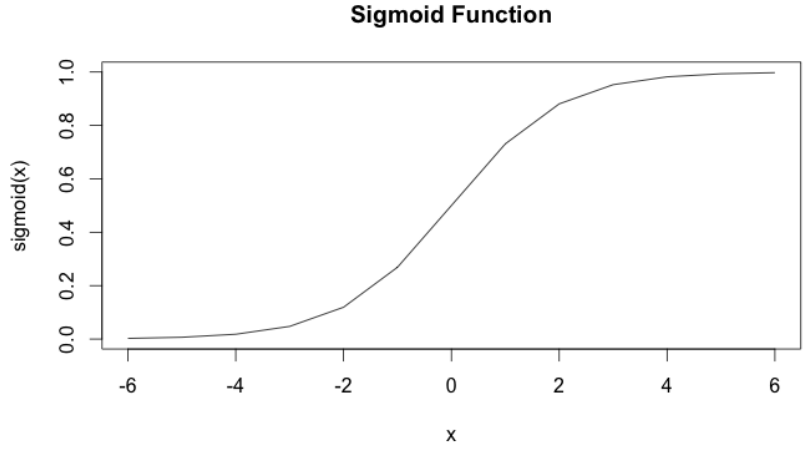

Figure 2.1: The sigmoid function, with x on the horizontal axis and *sigmoid(z)* on the vertical axis

Logistic regression is rarely used in the literature as a final model, although it is often used as a competitive model when evaluating others due its simplicity and interpretability. Khai-Wern *et al.* (2011) were able to apply logistic regression successfully for hazard analysis in the domain of landslide prediction.

A perceptron is alternate function-based classification model (Minsky and Papert, 1987). Like logistic regression, a perceptron is a linear classifier. There are conceptual similarities between logistic regression and perceptron models. Logistic regression can be said to combine the the input feature vector $X_{feature}$ with weights represented by Θ on a linear predictor function represented by the sigmoid function. A perceptron formalises this idea in its construction. A single node, representing an activation function, is provided with the input feature vector $X_{feature}$ for which each value x is multiplied by a specific feature weight w of the weight vector W. The activation function outputs a binary value of 0 or 1 if the dot product $w \cdot x$ is greater than zero. It is the weight vector W which needs to be fitted to a specific problem. Single perceptrons are strictly linear, and therefore are not suitable for problem spaces that would require a non-linear separation boundary. This limitation is overcome by connecting multiple perceptrons in successive layers, forming a multilayer perceptron. A multilayer perceptron is alternatively known as a neural network.

Neural networks are designed to mimic biological neurons. A neural network consists of interconnected nodes in successive layers, as can be seen in Figure 2.2. The first layer of the neural network assigns a node to each input.

Each node, or neuron, in the subsequent layers receives a weighted value from all of the nodes in the previous layer and sums them in a non-linear activation function. Neural networks can therefore be expressed as nested perceptron equations. The choice of activation function is consistent for each layer of nodes, and typically is the sigmoid function, or a function that displays the same output characteristics as the sigmoid function.

The weights between individual nodes are the parameters that need to be fitted for a specific problem. This is accomplished by applying the backpropagation algorithm to the model (Rumelhart *et al.*, 1985).

Figure 2.2 shows an example of a neural network receiving an input of a subset of $X_{features}$ used in this dissertation as nodes labelled I, one hidden layer of sigmoid function nodes H, and two output nodes which will respectively indicate the probability of the predicted 'safe' and 'unsafe' state of the rock as nodes O.

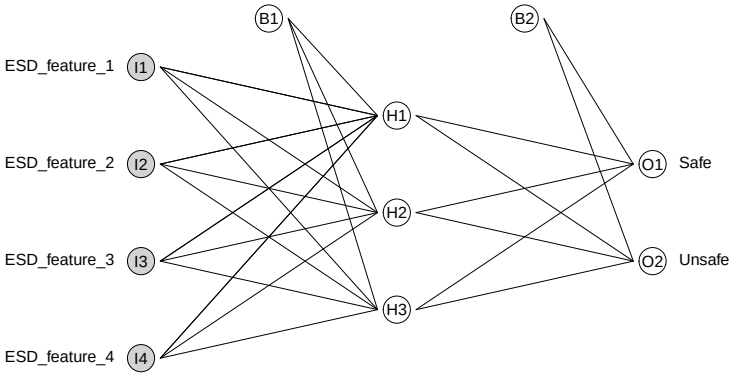

Figure 2.2: An example of a feed-forward neural network with one hidden layer

Neural networks are considered to be effective classification models, but offer limited to no interpretability of the underlying problem due to the emergent behaviour of combining the output of multiple non-linear functions. Interpretability is not a performance criterion for this project, however. McCrory *et al.* (2014) found that neural networks were as effective as any of the alternative models that they reviewed at predicting damage in carbon composites based on acoustic emissions. Aggelis (2011) performed *in situ* monitoring of fracturing concrete and succesfully used neural networks to predict the mode of fracturing based on the acoustic signature of the event. Similar research by

Kumar *et al.* (2013) showed the effective use of neural networks in the determination of rock structural stability based on the acoustic waveforms generated by drilling.

2.4.4 Kernel classification

Kernel-based classifiers are concerned with mapping data into a higher-dimension vector space where distinction between samples with separate classes is easier to model. This approach bears similarity to the instance-based classifiers already discussed in that a 'memory' of training samples is built up, and a similarity measure between a new sample point and the known training points provides the measure of classification.

The classification models that rely extensively on the kernel approach are commonly known as Support Vector Machines (SVMs). SVMs are a useful classification approach that has been successfully applied to problems where classification was made on acoustic techniques (Dhanalakshmi *et al.*, 2009).

Kernel-based methods extend on this idea to define the similarity measure between a known sample and the new sample as a kernel function. The kernel function is so chosen as to significantly increase the dimensionality of the space in which the comparison occurs. A linear separation boundary in higher-order space is potentially easier to find. The innovation in the kernel approach is that by choosing the right kernel, the so-called 'kernel-trick' is used to avoid the explicit mapping to higher-order space. This approach is able to provide the outcome of the similarity measure in the original space.

SVM classifiers are extensive in range, with novel implementations existing for specific problems in all domains. The popularity of SVM classifiers is due to the excellent results obtained with no significant increase in computational power due to the use of the kernel-trick. The choice of kernel function used in the SVM classifier defines the model. The first kernel evaluated in this comparison is the radial basis function (RBF) kernel. The RBF kernel is alternatively called the Gaussian kernel, due to its similarity to the squared Euclidean distance between samples, as can be seen in its formula:

$$K(x, x') = \exp\left(-\frac{||x - x'||_2^2}{2\sigma^2}\right) \tag{2.4.6}$$

A different kernel popularly in use for SVM classifiers is called the polynomial kernel. This kernel is based on the formula commonly also used in regression classifiers, such as logistic regression discussed in the previous subsection, and is given by

$$K(x, y) = (x^T y + C)^d \tag{2.4.7}$$

In this kernel, x is a subset of the training samples known as the support vectors, and y the testing sample to be predicted. The order of the equation

is determined by degree d, as can be seen. Notice that a value of $d = 1$ would imply that the kernel is linear. C is the cost parameter, which is included to vary the influence of the higher-order terms versus the lower-order terms of the polynomial.

Wu *et al.* (2013) found that the SVM classification approach resulted in a model with good generalisation performance when predicting rock bursts in China's coal mines based on acoustic emission sensing. Similarly, Ntalampiras and Roveri (2013) showed that it was possible to use SVM models based on micro-acoustic input data as a monitoring system for rock-collapse forecasting.

2.5 Conclusion

The literature review of the domains covered by the research in this dissertation revealed the shortcomings in the current approach to automated entry-examination-related assessment of loose rocks. There are specific limitations to existing approaches regarding the requirement for physical coupling of the instrumentation with the rock mass, as well as the pre-defined frequency ranges that are assessed in determining the state of the rock.

Features were identified that are currently in use in the domain of acoustic waveform transformations. A wide range of acoustic features is discussed in the literature, and an evaluation of the effectiveness of these features, as well as their current use in the determining structural cohesiveness, provided a guide to which features to select for this research.

Classification models used in acoustic prediction applications were assessed in groupings of common modality. Each of the classification models evaluated was found to be deployed successfully in applications where the only feature input is acoustic-based. Particular attention was paid to classification models that have been shown to be effective in the acoustic determination of structural cohesion.

Chapter 3

Methodology

3.1 Introduction

The methodology of the research is presented in this chapter. In developing this methodology, it is important to keep in mind the goal of the project that this process should achieve. A classification system for the Electronic Sounding Device is needed that is able to predict the structural state of rock being tapped with a sounding bar. This prediction is made by a model on the ESD by analysing the acoustic waveforms recorded and deriving a classification for the state of the rock. This chapter explains the research process involved in creating this classification system.

An overview of the methodology is presented in overview in Figure 3.1 on the following page. The individual items in the shaded cells of Figure 3.1 are the broad steps in the methodology, and coincide with the chapters of this dissertation. High-level sub-tasks of each step are presented in the unshaded cells next to the main steps in Figure 3.1. Each step is outlined in the rest of this chapter, and the deliverables between the steps are defined in these subsections.

The two main research instruments in this study, namely the ESD and the statistical programming packages, are separately described. The ESD, due to it being central to this research as well as to the need to give a full technical explanation of the system, is described in a chapter dedicated to it, Chapter 4. The statistical modelling environment packages and implementation language are described in the last step of the methodology, Model Implementation, in Subsection 3.3.7 on page 46.

The methodology of this research forms the critical basis of this dissertation. The success of the classification model that is generated by this process is fully reliant on the design of the methodology, and this directly affects the usability of the ESD as a potential tool in the South African mining industry. The next section outlines the goals of the methodology, after which the full methodology is detailed in Subsection 3.3.

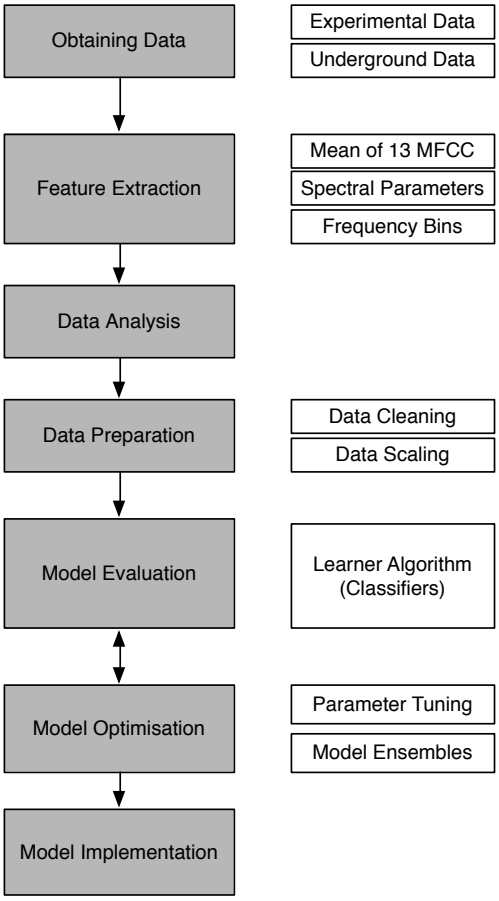

Figure 3.1: Methodology and structural layout of this research

3.2 Methodology goals

Classification models can relate input data to discrete predictions by generalising from example training data. This is a feasible method for determining problem phenomena when manual programming is not suitable for the task. It is however important to initially develop goals for the construction of a classifying system in order to provide focus in the multi-step construction process of classification models.

The design goals for this system are as follows:

Generalisation: The fundamental goal of the system is to be able to generalise beyond the training samples that it will 'see' during its construction. This goal should be considered in attempting not to overfit the classifying models onto the training data alone. To achieve this goal, experimental data and operational data will be collected and split into training and testing sets.

Representation: The choice of classification model or ensemble of models should be suited for the purpose. This goal implies that the suitable models should be identified in the methodology. It will be addressed in the learner-algorithm approach, described in Subsection 3.3.5.

Real-world applicability: The system should be tested for variance and bias as part of its testing. The limitation of false negatives, i.e. where the ESD predicts a stable rock when in fact the rock is loose, becomes the main performance evaluation characteristic of the classifying system.

Removal of researcher bias: With evaluation data in high dimensions it is important not to make assumptions without thorough analysis. In other words, intuition fails in high dimensions, and rigorous unbiased analysis is important before assumptions are made about the data or results. This goal is addressed primarily in the Data Analysis step of the methodology described in Subsection 3.3.3.

Wide coverage of models: The best classifying model for this problem can only be found if a wide range of models representative of most current approaches to classifying models is trained and evaluated.

These goals define the intent of the justification for the methodology described in the following section. The accomplishment of these goals is intended to deliver a classification system that is best suited for the ESD and its purpose.

3.3 Methodology outline

3.3.1 Obtaining data

In the research discipline of statistical classification, the quality of analysis and model prediction is directly dependent on the primary data on which it is based. This is the first challenge that this methodology needs to address comprehensively - it informs the validity of the conclusions and the applicability of the final system in an operational environment. This subsection provides an overview of the process of identifying suitable primary data for this project, and the plan of how to acquire it.

Ideal data

Understanding what an ideal dataset would look like for the purpose of training a classification model is the first step in specifying how the data collection process should be done.

Theoretically in an ideal dataset the input data would,

- be complete with no missing values
- have little to no noise
- contain no outliers
- be fully descriptive of the physical phenomena.

Furthermore, the dataset would also not be mislabelled, i.e. where any of the desired states is misleading or incorrect. This ideal dataset is of a theoretical nature and would be impossible to obtain from a true underground working environment where there are sources of constant noise, operational difficulties in collecting data are strongly prevalent and the structural nature of the rocks being tested is not fully known.

The strength of the idea of an ideal dataset is that it provides a theoretical underpinning against which to measure the realistic data obtained. Broadly, the definition of an ideal dataset for a specific classification problem is subjective and specific to the problem. Defining it is an iterative process as the problem becomes better understood. For the purpose of this research, the ideal dataset will be specified in Chapter 5. The specification of an ideal dataset can then be compared against realistically obtainable data in order to identify limitations and problems in the realistic dataset.

Data collection tool

Realistic and representative datasets need to be obtained as a prerequisite step before the development of any classification models can be attempted. The purpose of this research is to develop such classification models for operation on the ESD. It is therefore an intuitive and pragmatic decision to use the ESD platform to collect the training data sets. The ESD collects acoustic samples of the same size with a single consistent sample rate.

The ESD as a sensor platform is central to all facets of this research and therefore a separate chapter is allocated to describe it fully. Chapter 4 gives an in-depth description of the ESD's specifications and functional abilities.

Experimental data

Two datasets were collected for this research. The first was a set of acoustic waveforms collected in test conditions with an experimental test rig, where the physical variables that determine the input data and desired states can be fully controlled. This dataset will be closest to the ideal theoretical data specification and will be useful in validating the operational data in the next methodology steps. The full experimental set-up and the data obtained are described in Chapter 5, Section 5.3.

Operational data

For the purposes of this research, the most important dataset to train and test the classification models will be one that is was obtained from the working environment in which the ESD was deployed. The second dataset collected is a representative collection of acoustic data obtained by experienced miners in underground settings.

The location of the data collection encompasses different working areas with varying rock conditions. The data were also collected at various times of the underground working shift and therefore different ambient noise conditions should have prevailed during the collection of the data. Furthermore, the data were collected by different members of an experienced team of miners or rock engineers, who then represented different viewpoints on structural prediction.

The detailed breakdown of the data collection process in an underground environment is provided in Chapter 5.

3.3.2 Feature extraction

The identification and extraction of suitable input feature vectors for classification models are essential parts of any methodology regarding predictive classification. In summary, the feature extraction process involves simplifying the amount of data and resources required to describe a set of data accurately. When input data to a predictive model are considered too large to be processed, then it needs to be transformed into a reduced representational set of the features. In this project, the raw waveforms collected in the previous step of the methodology are too large to be used effectively as input data to the models due to the processing constraints on the ESD platform and the inadvisability of presenting unrefined data to classification models. This section describes the transformation of the raw audio waveforms into an effective input feature vector representation.

The difference between training a successfully predictive model and an unsuccessful approach is very much determined by the choice of features used. If many independent features can be found that each correlate well with the desired class, learning is potentially easy. If, however, the class is a very complex function of features, it might not be able to encapsulate the relationship in a classification model. The choice of features can often be where most of the novelty in creating predictive models lies.

The choice of features that sufficiently describe the problem phenomenon usually involves a novel approach, as opposed to a standard approach. To elaborate, each problem domain in classification has standard feature selections that has been proven to work well, but usually the choice of these features needs to be adapted to the specific problem. In the case of this research, features needed to be chosen that could comprehensively describe the acoustic waveforms generated by impact events. The choice of features therefore needs

to be focused on acoustic analysis, but this is a broad domain and therefore the relevant features of this problem need to be identified. In the literature review chapter of this dissertation, in Section 2.3, the features that were used in this problem case are described in detail. These features were found to work well with the acoustic analysis of impact events, such as a sounding bar striking a rock.

The feature set that will was used for this research is described in the following three subsections.

Mean of 16 MFCC

The Mel Frequency Cepstral Coefficient (MFCC) description of audio signals is frequently used in Human Language Recognition, the main research area in the acoustic-related classification domain. It is a useful technique for exposing the formation characteristics of an acoustic waveform. MFCCs have been found to be useful in the problem domain of describing impact events as well, as shown by research done in classifying percussion sounds such as drums being struck (see Section 2.3).

The MFCC of the raw acoustic waveforms were calculated. Usually, in the application of the MFCC, the extraction is done over a sub-frame of the larger acoustic waveform. However, the acoustic waveforms collected in this research are of sufficiently small duration, 190 ms, for the MFCC to be collected over the entire waveform. The mean of each of the 16 MFCCs is collected by averaging the coefficients over the impact duration, and is then used as a feature set for the description of the event.

A full description of the calculation steps followed in deriving the MFCC values for this project is given in Chapter 6, Section 6.5 on page 83.

Spectral shape parameters

What a miner hears when he or she taps the rock with a pinch bar is a response that would be described as ‘hollow’ or ‘solid’ by the miner. These descriptions correlate directly to the frequency distribution of the sound heard, where a lower frequency would sound hollow and a higher frequency solid. This suggests that a description of the frequency distribution would be useful as a feature set. The statistical shape parameters of a distribution are well defined, and it is these descriptors that were used as an additional feature set to describe the acoustic data.

The first four order moments of a distribution are:

- the spectral centroïd (or mean), $S_c = \mu_1$
- the spectral width (or variance), $S_w = \sqrt{\mu_2 - \mu_1^2}$
- the spectral asymmetry (or skewness), $S_a = \frac{2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3}{(S_w)^3}$

- the spectral flatness (or kurtosis), $S_f = \frac{-3\mu_1^4+6\mu_1\mu_2-4\mu_1\mu_3+\mu_4}{(S_w)^4} + 3$

In these definitions the spectral moment of order i is given by $\mu_i = \frac{\sum_{k=0}^{N-1} k^i A(k)}{\sum_{k=0}^{N-1} A(k)}$ where $A(k)$ is the amplitude of the k^{th} component of the Fourier transform of the acoustic waveform.

Frequency band content

Inspection of the frequency spectrum of the waveforms suggests that the distribution of the waveforms' energy over the frequency domain differs distinctly between individual samples. A feature set derived from aggregate log-energy values over certain frequency bands has been shown to be useful in describing the spectral characteristics of acoustic signals in discrete parameters, as was shown in Section 2.3.3 of the literature survey chapter.

Existing defined frequency scales, such as the Bark scale and the Mel scale, can be used to define candidate frequency bands. However, the earlier discussion in this section around engineering features particular to an individual problem suggests that features chosen from the researcher's understanding of the data can lead to a more suitable feature set. With this as an objective, empirical inspection of the collected data led to the identification of frequency bands that would be useful as reference bands for this research. These bands were then used as an alternative to the Bark Scale and Mel Scale.

The full process of feature extraction and a description of the final feature matrix is given in Chapter 6 and the reference bands used in this study is presented in Table 6.1 on page 80.

3.3.3 Data analysis

After the acoustic samples obtained for this research had been organized into input feature sets, it was important to inspect the data before presenting it to the classification model learners. An initial insight into the data benefits the research in a multitude of ways. The methods used to analyse the data, and the benefits that the research derives from this analysis, are discussed in this section.

Data analysis had to be done on the initial time-based acoustic waveforms, as well as the reduced-dimension feature set derived in the previous section. The evaluation of the unprocessed time-domain based acoustic waveforms was helpful in identifying for the researcher what an intuitively normal sample waveform would look like depicted as a time-series plot. This would also assist in the detection of anomalies during the data cleaning process that formed part of the next methodology step, namely the data preparation step.

The input feature parameters, extracted in the previous step of the methodology, were derived entirely from the frequency representation of the acoustic

samples. Inspecting this spectrum, and the features derived from it, is essential in gaining an understanding of the characteristics of the data and how they potentially relate to the prediction of structural stability in rock.

The visual inspection methods that were used to provide insight into the data are as follows:

- Time-series plots of the acoustic waveforms
- Power spectrum plots of the acoustic waveforms
- Box plots of each of the attributes in the feature set
- Histogram plots of each attribute in the feature set, divided into class values
- Pairwise scatter plots of all the attributes in the feature set.

These visual inspection methods provide an insight into the data which forms the basis for the following step in the methodology process namely the data preparation step. By understanding each feature attribute's internal composition and potential relation to other feature and class attributes, informed decisions can be made as to how to prepare and present the data in the learner algorithms in the model training phase of this project. The full process of data analysis and its conclusions are presented in Chapter 6.

3.3.4 Data preparation

The data preparation step in the methodology involves the additional transformation and removal of feature attributes and training samples. This is a necessary and important step before the features are used as training and testing sets in the creation of classification models.

Outlier detection and data cleaning

As can be expected, data samples collected from real-world environments by external experts will contain samples that are not related to the phenomena to be predicted. For instance, in the collection of acoustic samples for this project, rock engineers collected some samples in mining stopes during working shifts. Due to the difficulty of navigating these stopes with equipment in hand, erroneous readings were sometimes collected. The causes and sources of false readings are varied, and therefore this research will not aim to quantify them but rather accept that they are in the collected dataset and need to be identified through empirical means.

Two methods are available for sample cleaning. First, visual inspection of the collected waveforms is useful in identifying anomalies. An example of this

would be a waveforms that is completely saturated over its entire length, indicating either a very large continuous event - perhaps an operator accidentally brushing against the microphone of the ESD.

A second method of identification is the process of outlier detection. An outlier in this context is an observation in a dataset that appears to be inconsistent with the rest of that data set. Before outliers can be identified, it is necessary to determine what normal expected data would look like, and what would be considered anomalous. This is where the data collected from a controlled environment were used as reference.

The feature set of this problem contains descriptive parameters for each acoustic sample. These attributes are therefore good proxies for the samples in the determination of outliers. Outlier detection was done through three approaches:

- Extreme value analysis
  This analysis visualizes the data by using scatter plots, histograms and box and whisper plots to identify extreme values that are more than 2 or 3 standard deviations from the mean.

- Proximity methods
  Clustering methods are used to identify natural clusters in the data. A percentage distance from the cluster centroids is used and features outside this distance are filtered out.

- Projection methods
  These methods are useful for summarising the data in two dimensions. The observer then identifies the outliers by looking for anomalies.

Data scaling

The feature data at this point of the methodology vary widely in terms of range and distribution. It is worth remembering that the features are the result of chosen transformations of the original acoustic data, and that the representation of the features is therefore not rigid and can be adjusted to suit the purpose of this research better.

Feature scaling is a useful technique where the range of each input feature vector $X_{feature}$ is scaled to a certain uniform distance, typically within ranges of $X_{norm} = [0, 1]$ (normalisation) or $X_{stand} = [-1, 1]$ (standardisation, or obtaining zero-mean and unit-variance). The benefit of this approach is that each input feature vector provides a uniform distance range for classification models that are based on the concept of Euclidean distance between data points. Many models in fact require that the feature data have zero-mean and unit-variance, and will behave badly or not at all if this requirement is not met.

Normalisation works well with classification models that rely on the magnitude of features, such as k-means clustering and neural networks. Standardisation is essential for methods that train and predict on the distribution of feature vectors, such as SVM with an RBF kernel.

Due to implementation differences between models, and the surprising results that could be gained by approaching each model with varying feature sets, both standardisation and normalisation were used in this research for all the models during training. This can quickly highlight which scaling method is best suited for each model.

After the steps of data cleaning and data scaling have taken place, the input feature matrix is ready for use in training and testing models in the learner algorithm. The full description of the data preparation process is given in Chapter 6.

3.3.5 Model evaluation

The main goal of this research was to find a classification model that is best suited for the application of structural prediction from acoustic noise. This step in the methodology focuses on sifting through the available families of statistical predictive models and finding a suitable range of candidates to achieve the research goal.

For this step, it is useful to define terms that will be helpful in expressing the components of the system. To train a model a 'learner' algorithm is designed which inputs a training subset of feature values and desired class values, and aims to output a successful model. In this dissertation, 'learner' as a term is used interchangeably with learner algorithm. The learner algorithm should contain suitable candidate models that are trained and then evaluated against each other according to predefined performance criteria. A more intuitive description might even refer to a learner as a 'test harness' to spot-check models.

This step in the methodology was focused on the creation and running of the learner in order to find suitable models for this research.

The benefits of using a learner-based approach to evaluate several models can be summarised as follows:

- **Results:** A learner based approach delivers usable classification models faster than an unstructured approach. The potential failure of the learner may indicate that the data collected or the feature representation chosen in the previous methodology steps does not expose enough structure for any mainstream model to do well. A learner-based approach presents the results the researcher needs to move forward and optimise a given model, or backward to assess the results of initial steps in the methodology.

- **Objectivity:** There is a tendency on the part of the researcher to choose models that have worked well on past problems. A learner-based ap-

proach avoids this by automatically and objectively discovering the models that are best at picking out the structure in a particular problem.

- **Speed:** A lot of time could be spent on optimising models only to discover that a different family of models performs better on the research problem. The choice of a comprehensive suite of models in a learner avoids this.

To realise these benefits, an essential part of the formulation of the learner set-up is to include a representative model from all the classification algorithm families. A good mix of algorithm types will insure that no major classifying approach is left out during initial testing, enforcing the validity of choice of final classifies.

Each model under consideration in the learner should be given a chance to perform well. This implies that during the setup of the learner, each algorithm is given a fair chance by using experiments and heuristics to determine the variant that might best fit the problem.

Evaluation method

Evaluation of the performance of the various models needs to be done in a consistent, repeatable and transparent fashion for the choice of model to be valid. The evaluation component of the learner is discussed in this subsection.

The fundamental goal in this process is for the models to be able to generalise beyond the examples in the feature set. This is due to it being very unlikely that the chosen model implemented on the ESD will ever see those exact samples again once its in a working mine environment. It is easy for a classification model to appear to be doing well on a training set as it just needs to memorise the examples. A possible solution to this is to split the feature set into two subsets - a training set and a testing set. The model is trained on the training set without exposure to the testing set, and then after training it is judged on its performance of classifying the values in the testing set.

A more sophisticated approach to using a manually chosen training and testing set is called 'cross-validation'. This involves splitting the feature set into a number of equally sized groups of instances called 'folds'. The model is then trained on all folds except one that is left out, and the prepared model is then tested on that left-out fold. The process is repeated so that each fold gets an opportunity to be left out and used as the test dataset. Finally, the performance measures are averaged across folds to estimate the capability of the algorithm on the problem.

A performance measure that is relevant to this approach needs to be chosen from the established model performance measures. The choice of performance measure is based on the problem that the classification model needs to address. For instance, in this research the ESD needs to classify loose rocks.Therefore a performance measure needs to be specified that relates to the accuracy of the

ESD's detection rate, as well as to include a measure of how many mislabelled predictions are generated by the ESD. The performance measure chosen for this research, the area under the Receiver Operating Characteristic (ROC) curve, and the justification for using it, are detailed in Chapter 7, Subsection 7.2.2.

Classification models

The models that will be evaluated by the learner algorithm are listed in this subsection. As previously discussed, the range of models was chosen in order to maximise the benefit of the learner algorithm approach, which is to provide a representation from all the model groups.

Table 3.1 shows the models that were trained and evaluated in the learner algorithm, the model group representations which the models are associated with, and the model-specific tuning parameters that were used.

The tuning parameters of the classification models were determined by a grid-search method incorporated in the learner training approach. This approach meant models were trained iteratively with values obtained through pre-defined steps in the ranges of possible tuning values. These ranges were empirically determined or suggested by the domain literature. This concept is expanded on in the discussion of the next methodology step.

By means of the learner algorithm approach, the identification of models that predict the relationship between the acoustic samples collected and the structural state of the rock being tested can take place. The implementation of the learner algorithm and the models is detailed in Chapter 7.

3.3.6 Model optimisation

The temptation exists to implement the classification model that was found in the previous step of the methodology, and use it as the final result of this research. To do this, though, would mean missing out in the opportunity of model optimisation that could potentially lead to a significant improvement in

Table 3.1: Overview of classification models, their related family groups, and the model-specific parameters that were used in the learner algorithm

Grouping	Model	Model Parameters
Instance-based	k-Nearest Neighbours	$k \in [2\text{–}21]$
	Self Organizing Map	$xdim, ydim, xweight \in [1\text{–}10]$
Decision Trees	CART	$C_p \in [0\text{–}0.1]$
	Random Forest	$trees \in [250\text{–}2000], interaction.depth \in [1\text{–}5]$
Function	Logistic Regression	$\lambda \in [10^{-4}\text{–}10^{-1}]$
	Backprop-NN	hidden layer nodes $\in [5\text{–}20]$
Kernel	SVM - RBF	$\sigma \in [0.01\text{–}0.03], C \in [0.25\text{–}64]$
	SVM - Polynomial	$d \in [1\text{–}5], C \in [0.25, 0.5, 1.0]$

prediction results. This step of the methodology aims to analyse the models identified by the learner-algorithm of the previous step.

Model parameter tuning

Model tuning is an essential step in the classification model implementation. Models are parametrised so that their behaviour and characteristics can be tuned for a given problem. Some models do not have any parameters, and are considered optimised already. Most models do have parameters though, and the process of optimising these parameters suggests the search-nature of the problem. Therefore, finding the best combination of parameters can be treated as a search problem.

On the basis of this being a search problem, different strategies can be implemented to find a robust set of parameters for a given problem. Two standard techniques used in the field of model optimization are *grid search* and *random search*.

Grid search is considered the standard approach to model parameter optimisation (Bergstra and Bengio, 2012). The implementation of the grid search method starts with specifying the bounds of parameters to be considered. The grid search method applied to an SVM model with an RBF kernel would have to consider the two pertinent parameters of the model, namely regularisation constant C and kernel parameter σ. In the initial setup of the grid search method, a discrete set of possible values is chosen for each of the parameters. The method then trains the SVM model with each value of (C, σ) in the chosen set and evaluates the model's performance by means of the cross-validation process described in the previous section. The grid search then reports which combination of values showed the best performance by the model.

Random search is an alternative to grid search, and effectively functions in the same way with the pertinent difference being the choice in the set of parameters to be considered. Random search requires only the range of parameters to be specified, and then selects a set from a random distribution (e.g. uniform) from within the range. This has the benefit of eliminating the bias of the researcher when choosing the parameter values to be evaluated, removing the chance of missing potentially valuable values for the parameters.

Feature dimensionality reduction

The so-called 'curse-of-dimensionality' is a major concern in the field of classification models. The expression was created by Bellman (1956) to refer to the fact that many models work well in low dimensions but become intractable in higher dimensions. Generalising a model becomes exponentially harder as the number of features describing an acoustic sample grows, because a fixed-size training set will cover a much smaller fraction of the possible input space.

Some classification models can also be negatively affected by the presence of features which are non-informative. Adding features of non-informative data to models and evaluating the resultant accuracy of prediction would result in added complexity to most models.

To judge which features are redundant, or even detrimental, to the effectiveness of the model a similar search algorithm can be deployed that was used in the previous section to determine the optimal model tuning parameters. Such methods conduct a search of the features to determined which produce the best results when provided as input to the model.

Two methods were evaluated to determine whether the reduction of the amount of features could improve the classification model's performance. These are respectively recursive feature elimination (RFE) and principal component analysis (PCA). These methods reduce the dimensionality of the feature space without neccesarily sacrificing the usable information content in the data. The implementation and results of these methods are shown in 7.4

3.3.7 Model implementation

The goal of this research was to identify a suitable classification model that will operate on the ESD platform and provide the ESD's intended functionality of identifying loose rocks by listening to acoustic impacts on the rocks. This step of the methodology aims to operationalise the successful models derived from the previous steps of the methodology.

The two primary considerations for the implementation of the final classification model are the software and hardware environments of the ESD platform. The technical specifications of the ESD platform is discussed in detail in Chapter 4. The relevant details of the specifications relating to this discussion will be reiterated in this subsection.

On a hardware level, the ESD has a PXA255 microcontroller clocked at 400 Mhz. Practically, this level of processing power is capable for most forms of pretrained models in a suitable processing window period. The time frame for processing an acoustic waveform and outputting a prediction should be less than half a second in order to be effective for the user. This requirement gives a processing window of 500 ms for the execution of the model with the input of a single sample, as the case will be during the ESD's operation. The chosen model's operating time was tested on the ESD in order to assess whether this requirement is met.

The software system on the ESD is a modified version of the Linux operating system. This operating system is provided with compile and execution tool-chains for all the mainstream programming languages. Therefore, there is no strict limitation on choice of implementation programming language. The prime consideration then becomes the ease and strength of implementing models in the chosen language. The candidate languages are `C`, `C++`, Java and Python due to the experience of the researcher in these languages. The

choice of programming language was dependent on the support of libraries that contain optimised code for the classification models assessed in this research.

Development tools

Prototyping the models and visualising the input feature sets are tasks that are better suited to a desktop environment. For this developmental step, as detailed in the previous steps of the methodology, the programming language and statistical resources also need to be determined. The choice in this environment is relatively large, with a few strong contenders being currently in use by the data scientist community.

The three development libraries and platforms that were used in this research are detailed in the next paragraphs.

scikit-learn is a machine learning toolset that has been developed for the Python programming language. The strength of this package lies in the quick prototyping and consistent use of similar function calls across classification models. The added advantage of being in the Python programming language is the additional Python packages that can be used to import acoustic data, extract features and visually inspect the data. In particular, the *pandas* package in Python is very useful for data encapsulation and description, and will be used extensively in the Data Analysis step of the methodology. The DataFrame data structure in *pandas* is also equivalent to the standard DataFrame structure in R, which makes data swapping between the two environments seamless. The use of Python as a scripting language further enables relatively easy portability of the code to the ESD platform, as the ESD can natively run Python code.

Weka (Waikato Environment for Knowledge Analysis) is a suite of tools in the Java programming language developed at the University of Waikato in New Zealand. The main benefit of using Weka in this research lies in the design of the user front-end which aims to replicate and guide steps that is similar to those used in this methodology. In particular, it is designed to implicitly create a learner algorithm, previously described in the Model Evaluation subsection of this chapter. This makes it a good tool for spot-checking models. Its main limitation is the lack of customisation and the portability of its libraries - the model implementations cannot be extended to a workable version on the ESD platform.

R is the de facto standard for statistical computing, and is used widely by statisticians and data scientists. It is based on, and replaced, a niche programming language called S. The main advantage of R over the other prototype options is its focus on user extensibility. Due to this focus, researchers often share their code via R packages, and quite often the newest models in the domain can be found as an R package.

Each of these prototyping environments offers strong advantages, and this research did not focus on using exclusively only one of these environments.

For each stage of the methodology a particular environment is best suited, but none of these three options consistently outperforms the others over all the steps of the methodology.

A full breakdown of the coding of the methodology steps and the final implementation of the chosen classification model is given in Chapter 7.

3.4 Conclusion

In this chapter the methodology of this research was discussed. In particular, each specific task in the methodology was described to the level of detail necessary for the execution of this research to be clear, and the results to be justified. The steps in the methodology also provides a framework for the following chapters in the dissertation. The next chapter provides a technical overview of the Electronic Sounding Device, and then the following chapters detail the implementation and result of the methodology.

Chapter 4

Electronic Sounding Device Specification

4.1 Introduction

The Electronic Sounding Device (ESD) is chosen as the primary research instrument for this project. The ESD is designed to act as both the recording device for the labelled inputs for use in training the classification models, and as the production platform on which the models were be tested on and eventually used. As will be further elaborated on in this section, the ESD is capable of both the data gathering and data processing required for this research. Furthermore, the ESD is designed as a prototype platform capable of further modification, which makes it ideal as a testing platform for the results of this research.

The ESD itself is considered to be an innovative design, and the question could be asked why the design itself is not the subject of this research. The ESD was designed previously by parties other than the author, and therefore its design is beyond the scope of the current research of the author. The state of the device prior to this research was considered non-functional due to only the hardware aspects of the device having been designed and implemented before the current research. This author was responsible for finishing the testing and debugging stages of the ESD design project, during which time the research objective for this dissertation was identified. The author considers himself an authority on the design of the device after extensive experience in debugging, testing and commercialising it.

The ESD is being developed for commercialisation in partnership with Draxin. The product has the name of Rock Integrity Monitoring System (RIMS) (Draxin, 2014). The final cost per unit is not yet available at the time of this writing.

4.2 Electronic Sounding Device design

The ESD can be broadly described as an audio input classification device. It was created by the CSIR, originally by researcher Teboho Nyareli, under a project funded by the Platmine Collaborative Research Programme (Platmine, 2009). The duration of the project was from April 2008 to March 2010. The goals that the ESD was designed to fulfil are given in this subsection.

A collaboration of Platmine and the CSIR identified the need for a process enhancement in the making-safe procedure performed by miners during the entry-examination phase of mining (Nyareli *et al.*, 2009). The following design goals was initially specified for the device:

- The device should be capable of remote detection of the acoustic spectrum generated by the sounding process during entry examination.
- The device should be capable of embedded signal processing and decision making.
- The device should be available at a relatively low cost for the purposes of mass deployment in the mining industry.
- The device should indicate the predicted hazard state through an audible indicator, such as an on-board buzzer. Furthermore, the device should also indicate the hazard state visually by means of external LEDs.
- The device should be suitable for underground use. This design goal encompasses the following specifications: the device should be ergonomically acceptable to the typical end-user working in an underground mine environment; it should be rugged enough to withstand the operational working conditions; it should be reliable; and it should be capable of a battery life that will last throughout multiple work shifts.

The hardware platform of the ESD was designed with these goals in mind. The block diagram layout of the hardware implementation of the ESD is given in Figure 4.1 on page 52. This diagram shows the main components of the ESD, and the main connections between the components. Not every component will be discussed in detail as not all of them are relevant to the purposes of the dissertation. The hardware components that are directly related to the acquisition and processing of acoustic data and the classification models are further discussed in this section.

As can be seen in Figure 4.1, the central component of the ESD is the Gumstix motherboard. Gumstix motherboards are single-board computers known as computer-on-module (COM) systems. A COM can be seen as a variant of a single-board computer. In essence, a COM lies between a full-up computer and a micro-controller in nature. The design is centered on a microprocessor with RAM, input/output controllers and all the other features needed to

be a functional computer on a single board. However, unlike other available variants of single-board computers, a COM lacks the standard connectors for any input/output peripherals to be attached directly to the board. Due to the modular design of these peripherals in the implementation of the ESD, the COM motherboard is connected to a baseboard of the CSIR's own design. The specific Gumstix COM model used in the ESD is the Basix model (Figure 4.2), which has an XScale PXA255 processor, 64 MB RAM and 16 MB on-board Flash. In addition, an extra 1 GB of memory is connected to the board for the loading of the ESD's main programs, as well as the storage of audio samples collected for this research.

The microphone that is used for audio input into the ESD is an omnidirectional, waterproofed two-wire microphone. The audio signal between the microphone and the Gumstix board is amplified by an external 36 dB preamplifier with digital gain control, controlled via the Gumstix I^2C bus. The signal is sampled at a resolution of 20 bits. The frequency range for the microphone is 80 Hz to 10 kHz, and the sensitivity to sound is rated at -87 dB at a 73 dB sound pressure level. The limits of the frequency range can be clearly be seen in the input and training data presented in Chapter 5. This input frequency range is less than the extent of full human hearing which is roughly in frequency range of 20 Hz to 20 kHz (Dittmar, 2011). It does, however, cover the sensitive part of human hearing, as was discussed in Subsection 2.3.3. The sound recording on the ESD is sampled on two data channels. This is a legacy artifact due to the sound-processing chip typically deployed in stereo recording applications. In the case of the ESD, one channel is discarded due to the mono-recording limitation of one microphone. The sampling rate of the ESD is 44 100 Hz.

The version of the ESD used in this project as the primary research instrument is packaged in a rugged enclosure for extensive use in adverse underground conditions. The enclosure design is waterproof, dust proof and able to withstand impacts.

4.3 Training version of the Electronic Sounding Device

For the training data acquisition phase of this project, a modified version of the basic ESD device was created. This device contains exactly the same hardware components as the unmodified ESD, with the addition of a three-button dongle attached via a data cable to the ESD at approximately arm's length distance. This dongle is derived from a generic gate-opening remote control, from which it was appropriated by modifying each of the buttons to close a circuit connected via the data cable that is connected to the Gumstix via the baseboard of the ESD. A photo and the layout of the dongle are shown

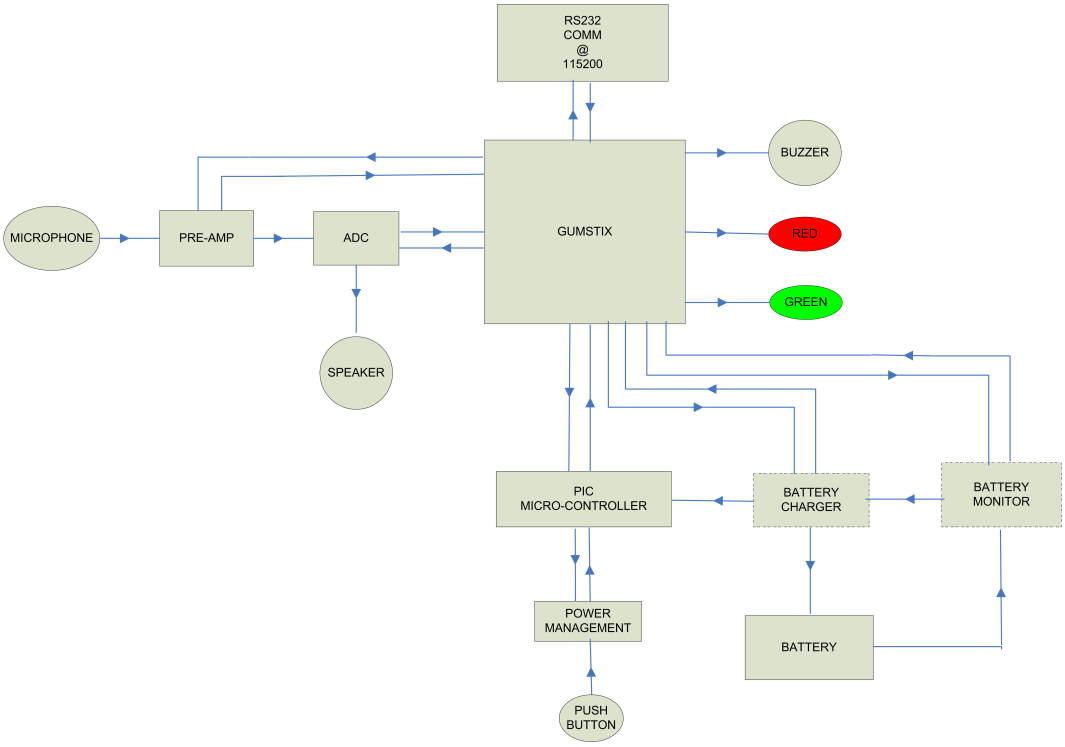

Figure 4.1: Block diagram showing the layout of the ESD hardware system

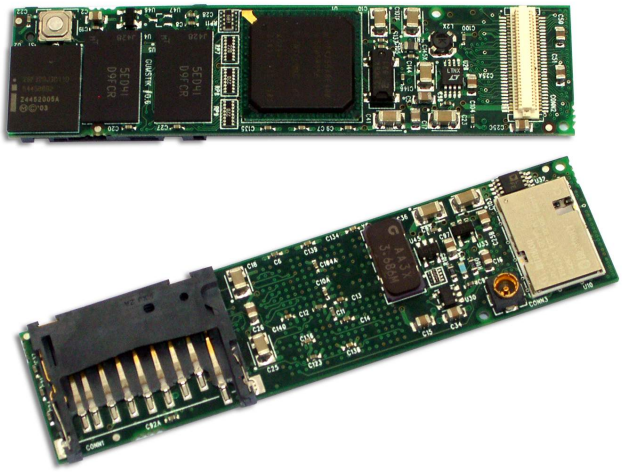

Figure 4.2: The Basix computer-on-module (COM)

Figure 4.3: Training dongle. This is an ESD attachment that enables a user to label training data during the recording process.

in Figure 4.3.

The training dongle has three buttons on it, coloured green, blue and red. Each of these buttons correspond to a label of the rock being sounded. The training version of the ESD listens for an impact sound in its environment, and records the acoustic waveform of the event. It then prompts the operator with an indicator on the dongle to provide a state label corresponding to the perceived structural safety of the rock being sounded. The green button records a 'safe' state, the red button an 'unsafe' state, and the blue button discards the recorded event. The device does not record any additional waveforms until a state has been provided.

This modification to the ESD allows for a quick recording and labelling of sounding events without requiring specialised skills by the operator. The feedback provided on the dongle itself, by means of its indicator, allows the operator to use the device without having to interact with the main ESD mounted on the hard hat. Due to the fact that the training ESD stops recording waveforms until a label has been provided enables the operator to put aside the dongle during the times when the process of training is not happening. The next chapter of this dissertation details the waveforms that are recorded during this process, and the use of this device in underground environments.

4.4 Conclusion

The use of the Electronic Sounding Device as the primary data collection tool and implementation platform for this research was discussed in this section. A model of the ESD that specialises in collecting training samples was specified as well. The data collected by this model of the device will form the basis of the next chapter.

Chapter 5

Acoustic Data Obtained

5.1 Introduction

In this chapter the acoustic waveforms collected by the ESD are detailed. Specifically, the specifications of the waveforms and the collection process involved are discussed in the sections of this chapter. Throughout this dissertation, the term 'raw acoustic waveforms' is used to distinguish the unprocessed waveforms collected with the Electronic Sounding Device, and this chapter focuses on aspects of these unprocessed waveforms.

The format of the raw acoustic waveform collected by the ESD is described first. Thereafter, the controlled environment from which the experimental set of primary data is collected is described, followed the full description of the collection process for the operational set of data recorded in working underground mining environments. The chapter concludes with the two sets of data - experimental and operational - intended for use in the next chapters of this dissertation.

5.2 Acoustic waveform specification

Understanding the feature vectors that will be used as input data to the classification system on the ESD depends on an understanding of the acoustic waveforms from which the input feature vectors are derived. This section details the composition of ESD-recorded waveforms.

Figure 5.1 provides the basis for the discussion in this section. It shows a typical waveform recorded on the ESD in an underground environment.

The event detection algorithm on the ESD is based on a simple moving-average algorithm. When in operational mode, the ESD continuously records an acoustic data stream from the microphone. Two average values are calculated and maintained at the arrival time of each new data point in this stream: the long-term average (LTA) and short-term average (STA). When a significant spike in the energy of the acoustic data stream occurs, the ratio

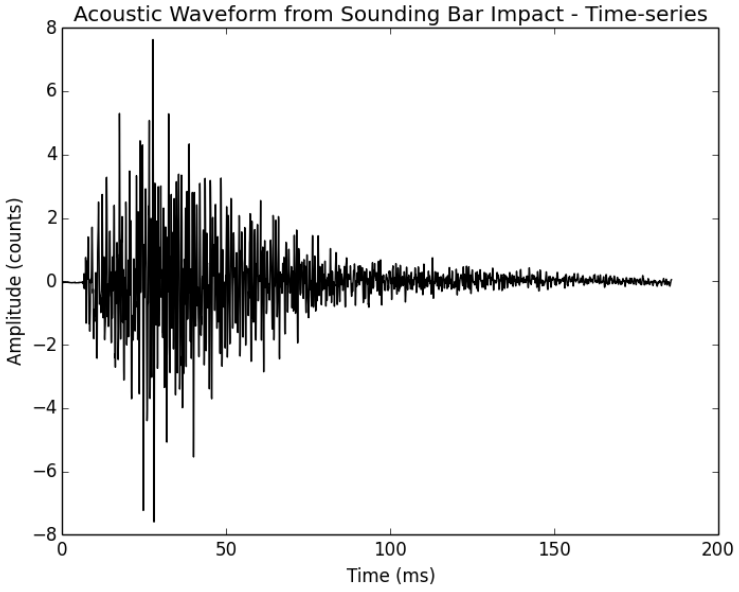

Figure 5.1: Example of an acoustic waveform collected by the ESD in operational underground conditions

of the STA to the LTA will rise significantly, and at a predetermined ratio in STA:LTA of 3:1 the ESD assumes that an impact event in the vicinity of the device has occurred. The data point at which this ratio is exceeded is considered to be the trigger data point. In Figure 5.1 the trigger data point can be found at the 13 ms mark. This corresponds to a spacing of 300 data points into the signal.

Note that the amplitude of the signal is expressed in arbitrary 'count' units. These counts are derived by the proprietary codec chipset in the recording chain of the ESD. This value also depends on the gains of the recording chain and the characteristics of the transducers. In practice, this ambiguity in the relation between signal energy and expressed units does not affect the distribution in the frequency domain of the signal. The frequency distribution of the signal is used to derive the input features for the classification models. The spectral distribution of this example signal can be seen in Figure 5.2 on the next page.

The choice of the event start location in the waveform should include a small period of the background acoustic noise before the event. In Figure 5.1, this is the quiet period roughly from 0 ms to 13 ms. This small period consists of 300 sampling points, and is included to provide an indication of the acoustic conditions in the area before the event was triggered. The event trigger is intended to be indiscriminate with regard to cause; it simply triggers when the STA:LTA predefined trigger ratio is exceeded. The inclusion of this initial period helps in the identification of the type of event that caused this

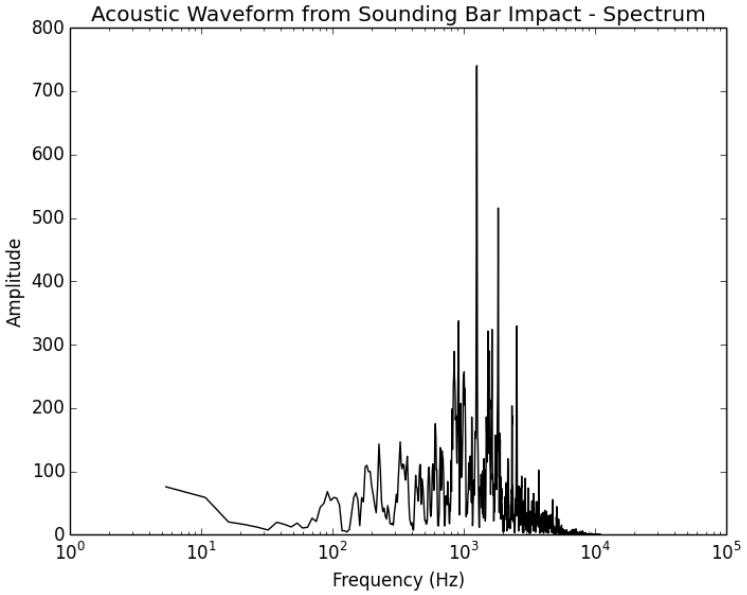

Figure 5.2: Periodogram of example acoustic waveform collected by the ESD in operational underground conditions

trigger. For instance, if an indication of an on-going event can be seen in the initial period of the waveform, then an assessment needs to be made as to whether the rest of the waveform does in fact indicate a sounding event recorded as opposed to another anomalous cause. This assists in the data cleaning process in the next step of the research methodology.

Most of the acoustic energy generated from the event can be seen in the waveform peak from 13 ms to approximately 80 ms in Figure 5.1. The expected energy peak of the impact event occurs soon after the triggered start of the event, and then subsides back to ambient noise levels after 100 ms. This expectation of the waveform's drop-off rate will also indicate during analysis whether a recorded noise is in fact from a rock being struck, as opposed to an environmental or anomalous noise that triggered the ESD.

5.3 Experimental set-up

The ESD is designed to be used in underground conditions and trained by experts on *in situ* examples of loose and solid rocks. A limitation on this research is that the rock examples that are demonstrated through this method by experts are not always proven actually to be structurally solid or loose, due to the fact that each rock mass is not subsequently barred down to test it. The examples are therefore expert assessments, but not objective fact. To mitigate this limitation in the initial development of the classifying algorithms, it is useful to have a controlled physical testing environment where the structural

states of the objects are fully known. This section describes the experimental test rig designed for the purposes of this research.

The intention of the test rig was to simulate a solidly attached rock and a loosely attached rock. To emulate this in a controlled environment a design of two separate blocks attached in various states of structural cohesion to a concrete roof was implemented. The schematic of the design can be seen in Figure 5.3. Component A is the design for the solidly attached block. This block is a solid marble cut-out that is solidly attached to the concrete roof by four iron bolts inserted at each corner of the block. Component B is the design for the loosely attached block. This block is also marble, but fractured in the middle and hanging from four iron bolts with a gap of 3 mm from the concrete roof.

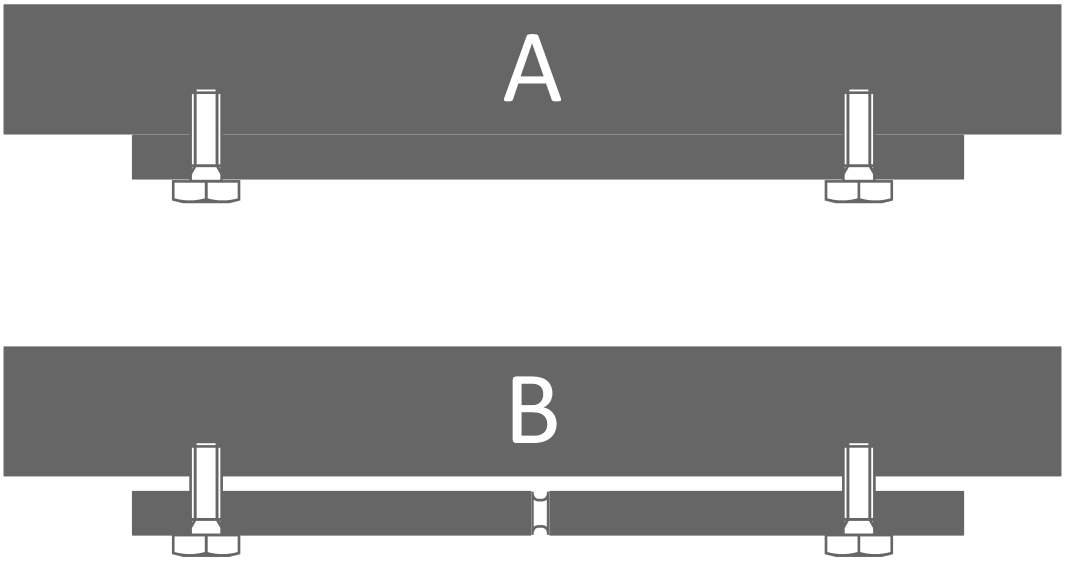

Figure 5.3: Experimental rig layout showing (A) the solid block, and (B) the loose and fractured block

A photo of Component A is shown in in Figure 5.4, and a photo of Component B in Figure 5.5. Note the indicators on the blocks marking the locations where the blocks are to be impacted. This is to ensure that most of the possible variations in the acoustic waveforms generated by the different areas of each block are consistently represented in the collected data set.

The aim of this experimental test rig was to enable the collection of an acoustic dataset that approximates the theoretical ideal dataset for this research. The attributes of the ideal acoustic dataset for this research that can be collected by means of the test rig must be as follows:

Figure 5.4: Experimental rig solid block

Figure 5.5: Experimental rig loose and fractured block

Fully descriptive: Each acoustic sample is accurately matched to a completely known structural state. There is no ambiguity in the accuracy of the state.

Distinguishable: Acoustic samples generated from the solid block should be distinguishable from acoustic samples generated by the loosely attached block. A clear hyperplane should exist between the two subsets of the data collected.

Repeatable: The size of the sample base should be easily expandable. Access to the experimental rig should be easy on an *ad hoc* basis if more samples are required.

Noise-free: The samples should be realistically noise free. During the process of collection, background noise sources are monitored and avoided. The acoustic waveforms recorded by the ESD are purely from the impact of the sounding bar on the marble blocks at predetermined areas on the marble blocks.

During the collection of acoustic samples from the test set it is noted that the samples from the two blocks sound very distinct to a human observer. This clear disparity in the acoustic profile from the two sets of samples will make it easier to classify waveforms by means of a classification system. Caution is expressed that this approach therefore does not lead to ambiguous samples, and that this characteristic of the test platform disqualifies it from being a direct proxy of an underground mining working area.

A total of 200 acoustic samples evenly split between the solid and loose configuration of blocks is collected as the control sample for this research. These samples are intended to be used as demonstrators of distinguishable data separation and feature correlation in Chapter 6, and as basic validators of potential models in Chapter 7.

5.4 Underground data collection

This section details the collection of the primary acoustic data used in this research for the training, testing and validation of the classification models. The acoustic data were collected in operational underground conditions by representatives from the Rock Engineering Department at Gold Fields Driefontein Mine. These data are intended to be representative of the types of acoustic data on which the ESD will be expected to base predictions in its functional deployment. The specific methodology and the results of the collection of the samples are discussed in this section.

Ideally, the ESD should be able to function in most of the environments that can reasonably be expected in a specific mining area. For instance, a

mining concern should be able to deploy the same ESD in a few mining shafts that are in close geographic proximity to each other. The problem with this scenario is that there are a few different environmental variables that are not consistent across shafts, or quite often not even within any particular shaft. The methodology used to obtain the labelled acoustic samples should aim to collect a representation of acoustic waveforms from a spread of locations that include all these variables. Therefore, a partnership between the CSIR and the Driefontein Gold Mine operation owned by Gold Fields was established to collect labelled acoustic waveforms from separate shafts across the mine.

Driefontein Gold Mine's geology includes three gold-bearing reefs, each with its own distinct characteristics. The relevance of the existence and characteristics of the reefs to this research is that additional descriptive data on the location of each collected acoustic sample can potentially be used in the data analysis part of the methodology and as additional descriptive feature data in the input feature vectors to the the classification model of the ESD system. The ore body cross-section is shown in Figure 5.6. The three reefs, in order of their relative distance from surface, are:

- the Ventersdorp Contact Reef (VCR) - bordered by two lava formations named the Westonaria formation and the Alberton formation
- the Middelvlei Reef (MR) - bordered by a quartzite formation
- and the Carbon Leader Reef (CL) - also bordered by a quartzite formation.

The data collection protocol was specified by the researcher. Samples were collected by Gold Fields staff in teams consisting of three people equipped

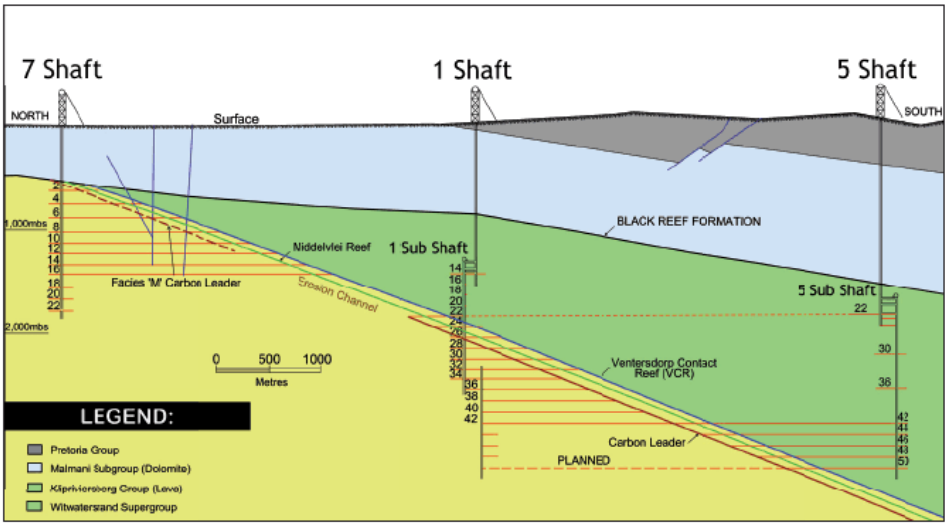

Figure 5.6: Typical cross-section through the Driefontein Mine ore body

with a training model of the ESD as described in Subsection 4.3. Each team was led by a Mine Engineer assisted by a Mine Technician and the designated entry-examinator of the area where the samples were collected. An on-site duration of one hour per collection process was chosen, to correspond with the existing entry-examination period of a working shift. Eight operational sites were selected for the collection of samples. Table 5.1 shows the planned stopes for labelled acoustic sample collection, as well as the reef that occurs at each stope. Each chosen stope is associated with a specific reef, and the combination of the geology of these stopes was intended to be representative of the geology that occurs at Driefontein Mine.

The collection protocol included the manual recording of the conditions observed at each collection site at the time of recording. Three variables were recorded as part of the description of each site - the rock fracturing condition, the prevalence of moisture in the working area, and the phase of the working shift during which the team made the recordings. These three variables were categorised into discrete states as part of the data collection protocol:

Ground conditions: Crushed, Fractured, Jointed, Disturbed, Intact.

Water conditions: Flowing, Dripping, Moist, Damp, Dry.

Part of shift: Entry Examination, Mid-shift, Special instruction.

The numerical labelling of the sample sets collected was done according to the date on which these sets were collected. For example, the third sample set was collected at Stope 38, Panel 25 at Shaft #1. This set was collected in the third data collection session, and it is therefore labelled Set 3.

During the collection of the Set 1 data, the operators had to switch ESD devices during the middle of the collection session due to the battery failing on the first device. The Set 1 data were therefore collected with two different sensors. To reflect this, each device's collected data are labelled separately, in order to maintain the distinction for later data analysis per set. Set 1 is therefore subdivided into Set 1A and Set 1B for the rest of the research.

A full description of each collected dataset is shown in Table 5.2. Note that the stope numbering in Table 5.1 and Table 5.2 is consistent for cross-lookup purposes.

5.5 Observed conditions at collection sites

The variety of conditions found in underground stopes essentially makes each of the collection sites unique. General conditions, such as the ground, water and time-of-shift, have already been described. In addition, special conditions existed at the time of collection at each of the sites, and these need to be mentioned in order to provide a full description of each of the sets of collected

Table 5.1: The specific stopes and reef types at Gold Fields Driefontein operation where the primary training data were collected

Shaft	Stope	Panel (Area)	Reef Type
1	38-25	9 West	Carbon Leader - Quartzite
2	24-29	8 and 9 West	
	16-38	1 East (Upper dip)	VCR - Westonaria Formation Lava
	20-28	4 West (Upper dip)	
	20-36	6A West	
8	8-36	2 West	VCR - Alberton Formation Lava
	8-45	9 West (Upper dip)	
	6-35	2 East	Middelvlei Reef - Quartzite
	10-35	3 West	

Table 5.2: Sample set numbering and amount of labelled usable samples collected per working area

Set	Stope	Samples	Ground	Water	Shift (time)
1A, 1B	20-28	84	Intact	Moist	Entry (8:00)
2	20-28	57	Intact	Moist	Entry (9:00)
3	38-25	42	Fractured	Moist	Entry
4	20-36	56	Crushed, fractured	Dry	Entry (20:00)
5	10-35	79	*Not recorded*	*Not recorded*	*Not recorded*
6	16-38	179	Fractured	Moist	Entry
7	8-36	78	Fractured	Dripping	Entry
8	8-45	84	Fractured	Dry	Entry
9	6-35	51	Intact	Dry	Entry

acoustic samples. This subsection briefly covers the conditions found at each site.

Stope 20-28, the site of Set 1 and Set 2, had very unstable hanging-wall conditions. During the collection of Set 2, in particular, the Gold Fields team did extensive barring down of loose rocks during the collection process. This resulted in the majority of readings from this area being of loose rocks.

Stope 28-25 was the collection site for Set 3 of the data. At this stope, most of the hanging wall was described as 'clearly loose' by the collection team. Very few acoustic samples of solidly attached rock were collected. Most of the rocks that generated the samples of loosely attached rocks were subsequently barred down. This provides special credence samples of this set being labelled 'unsafe' as empirical barring correlates with the readings.

The mining site at stope 20-26, where Set 4 was collected, had a special excavation protocol in place. The assumption at this stope was that the ground is loose and depends on continual support before and after blasting. Barring

down of loose rocks in this area does happen, but less often than in other areas because it is of limited effectiveness compared with the deployment of pervasive support in the area. Reduced follow-up barring after the collection of samples was due to the operational paradigm of this stope.

Set 5 was collected at stope 10-35. It was not possible to retrieve the records of the conditions at this stope. The samples collected at this stope are therefore not described to the extent of the other sets of collected acoustic data. The individual samples are labelled by the training version of the ESD as loose or solidly attached, and are still considered potentially useful as a set of data for this research. During the analysis of the data, special attention was given to this set to identify potentially erroneous readings.

During the collection of Set 6 at stope 16-38, part of the collection process occurred directly on the mining face of the stope. This means that the acoustic samples collected were not only from the hanging wall as usual, but also from the the face wall. Eighteen acoustic samples from the sounding process on the face wall were collected. These samples were kept in the set to represent a representation of sounding that can occur during a normal shift.

Set 7's samples were collected through the use of sounding bars of two different lengths. Stope 8-36 was excavated at different hanging wall heights, and during the first half of sounding the hanging wall was 2 meters above the heads of the collection team. A special extended sounding bar was used to collect these samples. The second half of the set was collected with the normal sounding bar.

Sets 8 and 9 were collected by Gold Fields teams while they were investigating two stopes that had recently had significant fall of ground (FOG) events. These events caused a significant portion of the hanging wall to collapse, and the remaining excavations therefore had a high remaining hanging wall. Another special condition at both these sites was that clean-up operations were in progress to remove the FOG detritus. The ambient noise of the clean-up process would affect the samples collected from these areas. In particular, the team made a note of the air hoses being used during clean-up in their vicinity, which falsely triggered the ESD at certain points during the collection process.

5.6 Conclusion

This chapter described the acoustic waveforms that form the basis of the primary data of this research as well as the two processes used to collect sets of these data for the subsequent phases of the project. The acoustic waveform were demonstrated by means of an example waveform collected in underground mining conditions. The ideal dataset was defined, and an experimental rig was specified and implemented for the collection of the ideal dataset. This set will be used as a control set in the subsequent phases of this project.

The purpose of this chapter was to describe the collected labelled acoustic datasets from underground mining environments that will function as the primary data for the next phases of this project. The underground collection protocol was given, and the various operational sites where the acoustic data were collected were described.

In total, 710 labelled acoustic waveforms were collected from underground operational mining environments by experts in the entry-examination field. In addition to the waveforms collected in typical operational conditions, 200 labelled samples were collected from the controlled experimental rig. These waveforms were transformed into functional feature vectors for the training and testing of classification models as explained in the next chapter of this dissertation.

Chapter 6

Feature Data Derived

6.1 Introduction

The acoustic data that were collected in the previous chapter cannot be applied directly to the ESD classification system without reducing the dimensionality of the raw data first. This process is called 'feature engineering', and is the focus of this chapter. Feature engineering is the process of transforming raw data into features that better represent the underlying problem for use in the predictive models, resulting in improved model accuracy on unseen data. This chapter will show the three chosen transformations that were applied to the acoustic data to generate a combined feature set for use by the models in the next chapter, as well as providing an insight into the collected dataset. Figure 6.1 shows the initial transformation steps to generate the complete feature set.

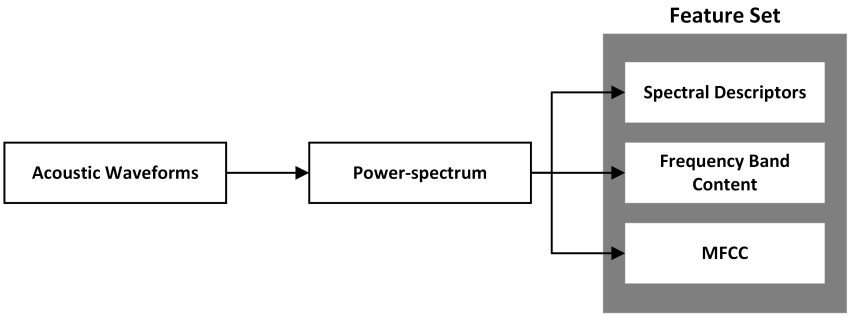

Figure 6.1: Process to generate the complete feature set

In the stated methodology of this research, three intermediate steps between the collection of the raw data and the creation of the models are required. These three steps are:

- the Feature Extraction step, where the features are derived from the raw data (see Section 3.3.2)
- the Data Analysis step, where the resulting features are analysed for useful information (see Section 3.3.3)
- the Data Preparation step, where the features are additionally transformed and cleaned (see Section 3.3.4).

These three steps of the methodology are not intended to occur linearly in the creation of a classification system: they form an iterative process in creating and transforming a feature set that enables the successful realisation of a classification system. In order to align closely with this iterative process, this chapter will detail these three methodology steps for each of the feature transformations.

Another component that is common to the analysis of each of the transformations is the distribution of the class values that the classification system will eventually aim to predict. A description and analysis of the class values are therefore provided before the subsections detailing the feature transformations themselves.

The outcome of this chapter's work is the provision of a combined feature set consisting of the results of each of the feature transformations. This combined feature set forms the basis of the data that will be used to train and test the classification models to validate their results.

6.2 Class values

The class values of the feature set are the values that the models should ideally be able to predict accurately if given the rest of the feature set. In terms of this research, these class values are the labels that describe whether an acoustic sample is generated by a loosely attached rock or a solidly attached rock. In the previous chapter an overview of the samples collected was given - in Section 5.4, and specifically Table 5.2 on page 63, the quantity of samples was presented. Before discussing the rest of the feature set, it is necessary to provide a deeper analysis of the class values to relate them to the rest of the feature set.

Each of the subsets of data collected was recorded in a specific area with specific conditions. Some of these working areas were able to provide sample acoustics of both solidly attached and loosely attached rocks, whereas in other areas only samples of loosely attached rock could be recorded. The breakdown of the samples collected for each set is given in Figure 6.2 on the next page. Note that the class values in the feature set are indicated as 'safe' or 'unsafe'. This is to indicate whether the sounded rock mass is considered safe or hazardous for miners in the area, due to its perceived cohesion to the larger rock

mass. This class labelling will be used consistently in the discussion of the dataset.

The control samples from the experimental rig described in Section 5.3 are evenly divided between the 'safe' and the 'unsafe' states. In the 200 samples collected from the experimental rig, the collection procedure resulted in 100 samples of each class.

In the following sections of this chapter, data cleaning operations are performed on the feature set based on the identification of anomalous observations in the data. The process reduces the quantity of usable samples in the complete feature set. The omission of certain samples might affect the distribution of class values as given in Figure 6.2.

6.3 Spectral descriptors

The spectral descriptors of the frequency distribution of the waveforms are defining features of the shape of the distribution, providing a description of the 'personality' of the distribution. This is particularly useful as a feature for this research since each descriptor provides a univariate value that sums up a characteristic of the distribution. This effectively reduces the dimensionality, or number of variables, that a model will need for insight into an acoustic sample. The spectral descriptors that are used as features for this research and

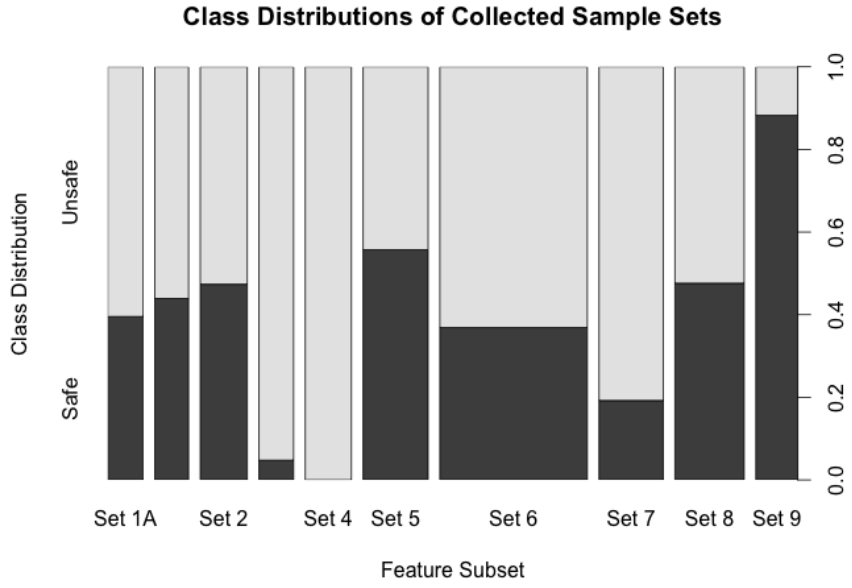

Figure 6.2: The relative distribution of class values over the collected subsets of labelled acoustic data. The vertical axis indicates the relative distribution of the class labels and the horizontal axis indicates the relative sample quantity per subset

are investigated in this section are the first four central statistical moments of the frequency distribution, as expressed in the power-spectrum form.

To recapitulate, the first four central moments are:

- the spectral centroïd (or mean)
  $S_c = \mu_1$
- the spectral width (or variance)
  $S_w = \sqrt{\mu_2 - \mu_1^2}$
- the spectral asymmetry (or skewness)
  $S_a = \frac{2(\mu_1)^3 - 3\mu_1\mu_2 + \mu_3}{(S_w)^3}$
- the spectral flatness (or kurtosis)
  $S_f = \frac{-3\mu_1^4 + 6\mu_1\mu_2 - 4\mu_1\mu_3 + \mu_4}{(S_w)^4} + 3$

In these definitions the spectral moment of order i is given by $\mu_i = \frac{\sum_{k=0}^{N-1} k^i A(k)}{\sum_{k=0}^{N-1} A(k)}$ where $A(k)$ is the amplitude of the k^{th} component of the Fourier transform of the acoustic waveform.

Each of these central moments is applied to the datasets in turn, and the resulting single-dimension vector corresponding to each central moment is then analysed in order to gain insight into the data. This analysis is given in for the following subsections.

6.3.1 Spectral descriptors - Experimental data

The acoustic waveforms collected from the test rig are analysed in this subsection to provide an insight into data that is noise-free and controlled. This test dataset will also establish the validity of the use of spectral features as part of the complete feature set. The frequency distribution of these samples is first derived through a standard Discrete Fourier Transform (DFT) from the time-domain acoustic signal into its frequency-domain representation. The spectral descriptors are then calculated from the frequency distribution and analysed for insight into the feature set.

The Tukey boxplots of the four spectral descriptors derived from the experimental data is given in Figure 6.3. The Tukey boxplot is a useful visualisation for these distributions. In a Tukey boxplot the box encompasses the distance from the first and third quartiles of the data, with the line in the box indicating the mean of the data distribution. The distance between these two quartiles are known as the interquartile range (IQR). The whisker length of the boxplot shows the lowest datum still within 1.5 IQR of the lower quartile, and the highest datum still within 1.5 IQR of the upper quartile. Any values identified as outliers on this boxplot should be evaluated for anomalous conditions, and considered to be cleaned from the dataset. In order to illustrate the structural

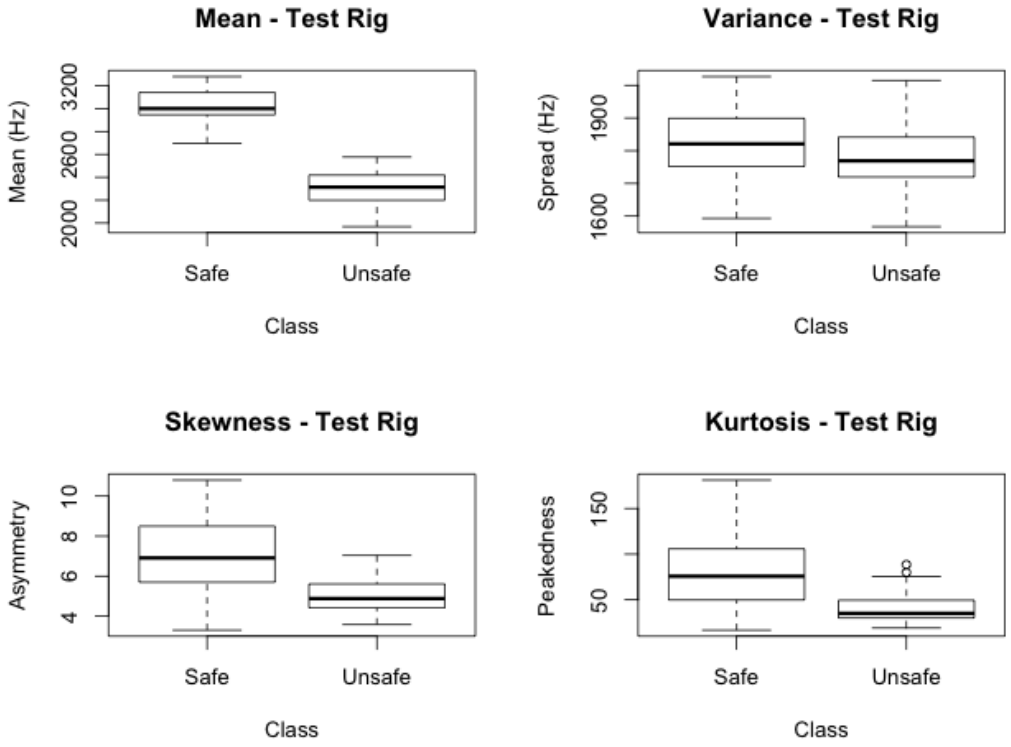

Figure 6.3: The distribution of the spectral descriptors by class values for the acoustic samples collected from the test rig

differences in the feature set, the boxplots of Figure 6.3 are categorised by class states on the horizontal axis.

The first moment, S_c, of the distribution corresponds to the mean of the distribution. The first boxplot of Figure 6.3 shows the distribution of the S_c values of the acoustic samples collected from the test rig. The visualisation of the spectral mean over the experimental dataset gives an indication of the combined weighted mean of the frequency distribution over the entire set. This provides a useful insight into the dataset, due to the mean of the frequency content of the waveforms being directly related to the perceived pitch of the acoustic sample. This has a bearing on the hypothesis that the distribution of the pitch of the sample can be used to make a predictive classification of the structural state of the rock. As can be seen, the distributions of the S_c value by class are clearly distinct for each class. There is no overlap of the two distributions. In fact, a model with a decision boundary at approximately 2 600 Hz would be able to classify the two classes completely. This is the advantage of using a controlled environment such as the test rig: visualisation of potential decision boundaries becomes possible to provide justification for using spectral parameters as feature vectors.

The second moment, S_w, indicates the variance of the frequency distri-

bution of the test set. There is not much that can be derived from visual inspection, apart from the similar ranges of two classes, since all of the sample's frequencies are concentrated between 1 600 Hz and 2 000 Hz from the mean. No outliers are present.

The third moment, S_a, indicates the skewness of the frequency distribution of the test set. As can be seen in the third boxplot of Figure 6.3, both the class distributions skew towards positive values. Essentially, both the distributions are skewed towards the higher frequency values. This provides a guideline to the evaluation of the S_a value of the operational set where the same skewness should be present.

The fourth moment, S_f, is a measure of the peakedness of the distribution. A large value for a sample would indicate a very strong dominant frequency peak existing in the frequency distribution. The last boxplot of Figure 6.3 indicates the distributions by class of the kurtosis values for the test rig set. As can be seen, relatively high kurtosis values are present for the 'safe' class. This would indicate that the block of the test rig corresponding to the 'safe' designation responds with comparatively high energy in a narrow band of frequencies. Most of the 'unsafe' samples are more evenly distributed over their frequency range, though two outliers can be seen in the boxplot. Visual inspection of the waveforms generating these outliers does not indicate anything anomalous with these particular samples, so they were not cleaned from the test data set.

The examination of the spectral descriptors for the experimental data collected from the test rig leads to insights into the frequency distributions that can be applied to the operational set.

6.3.2 Spectral descriptors - Operational data

The boxplots for the spectral descriptors for the full operational dataset collected from the Driefontein Gold Fields mine are shown in Figure 6.4 on the facing page. It is immediately noted that these distributions are significantly different from those of the spectral descriptors for the experimental data examined in the previous subsection. This subsection will analyse the distributions and perform appropriate data cleaning on the operational set.

The expected distribution of the S_c parameter, the spectral mean of the acoustic samples, over the operational dataset indicates that the 'safe' class's mean values are on average lower than those of the 'unsafe' class. This is not an expected result, as the perceived pitch for structurally secure rocks is typically higher than that for loose rocks. This result indicates that the dataset needs to be examined for inconsistencies and errors.

What can also be noticed in Figure 6.4 is that the boxplots for the distributions for the skewness (S_a) and the kurtosis (S_c) parameters indicate a significant number of outliers. When comparing the range of the parameters in these distributions with those of the experimental dataset a wide disparity is

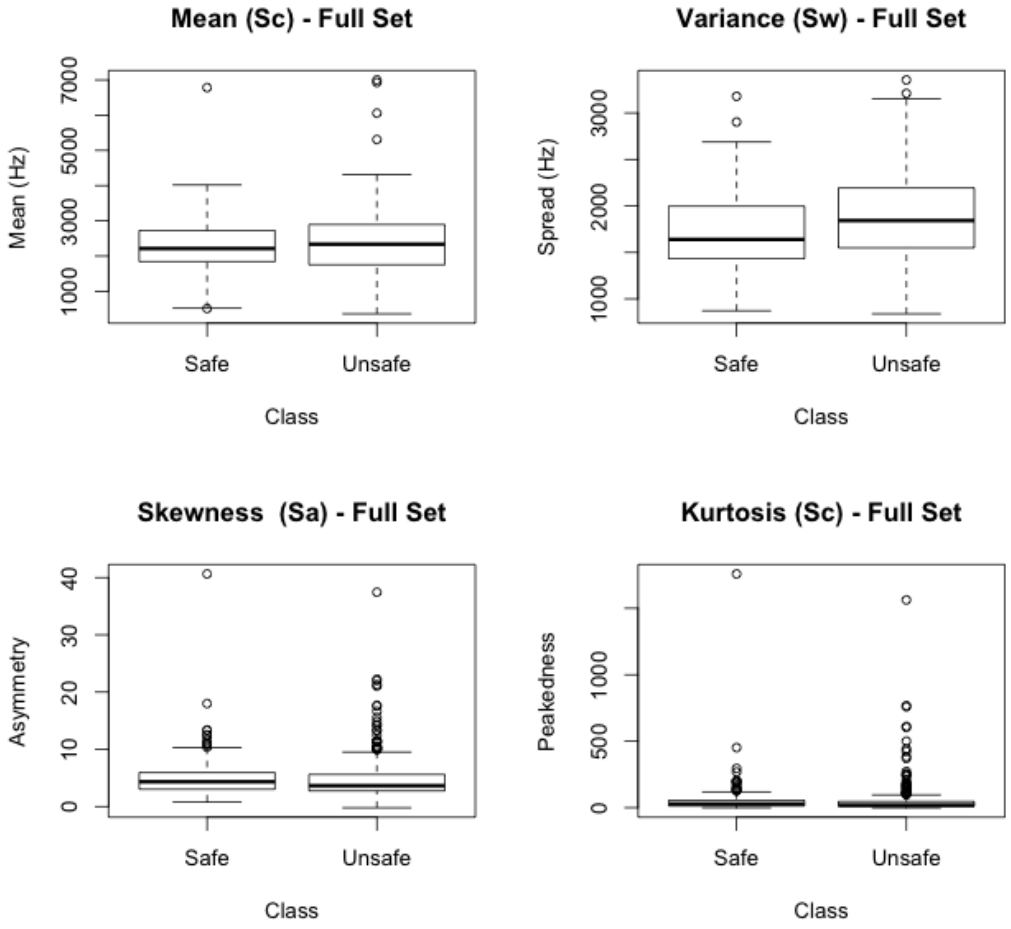

Figure 6.4: The distribution of the spectral descriptors by class values for all the acoustic samples collected from Driefontein Gold Fields mine

noticed. For the experimental dataset, the skewness parameter was bounded by 10, and the peakedness parameter is by 200. For the operational dataset, the skewness parameter peaks at 40 for both classes, and the peakedness parameter is bounded by nearly 2 000. This is an order of magnitude difference between the two sets, a disparity which needs to be addressed by thorough data examination.

In the methodology section of this dissertation, in Subsection 3.3.4, three approaches for data cleaning are given. To reiterate briefly, they are as follows:

1. Extreme value analysis
2. Proximity methods
3. Projection methods.

The first of the approaches is the elimination of samples through the method of extreme value analysis. This can be combined with the third approach, projection methods, by plotting the data visually. The combination of the two approaches works well with the statistical descriptors of the operational dataset due to the fact that the set is easily summarised in the boxplot figures which indicate outlier values. The presence of outliers is particularly true for the S_a and S_c parameters, as can be clearly be seen in the last two boxplots of Figure 6.4.

When performing data cleaning on a dataset, the approach should always be justified and the resulting exclusion criteria for discarded samples should be transparent. If samples that do not fit the preconception of what the data should be are discarded without insight, the process threatens to become 'data wrangling'. Data wrangling is akin to changing the data to suit the researcher's need, as opposed to leaving the data in a state where it describes the problem phenomenon completely. Data wrangling should be avoided in research in general, and in classification problems in particular. Keeping this in mind, the operational dataset's outlier samples were evaluated to gain an understanding of the samples that are indicated as outliers.

Visual inspection of the time-series plots of the outliers provides an indication of the type of waveform that is considered anomalous. Three types of anomalous waveforms are found in the operational dataset. An example of each type is shown in Figure 6.5 on the next page and they can be described as follows:

Saturated: These waveforms are triggered by the microphone of the ESD physically brushing against something. Typically, this can happen if the operator of the ESD is handling the unit and brushes a finger over the microphone. This results in a completely saturated waveform. These waveforms are discarded.

Impulse: These waveforms are triggered by what seems to be a single impulse event. No other acoustic information is present before of after the impulse event. Exactly the same event occurs periodically over all the subsets of the operational data. The characteristic of the waveform, with a high peak, shallow depth and no subsequent taper-down effect, is identical each time it occurs. This suggests that this type of waveform is artificial, and is generated by the sensor in the ESD. It was not possible to replicate this in the experimental test rig, and therefore it is assumed to be generated by a malfunction in the ESD sensor-chain in strenuous underground conditions.

Low-frequency: Waveforms with this characteristic display a low-frequency element that dominates the entire waveform. The ESD is easily triggered in this state, as the short-term ration in the STA:LTA triggering ratio

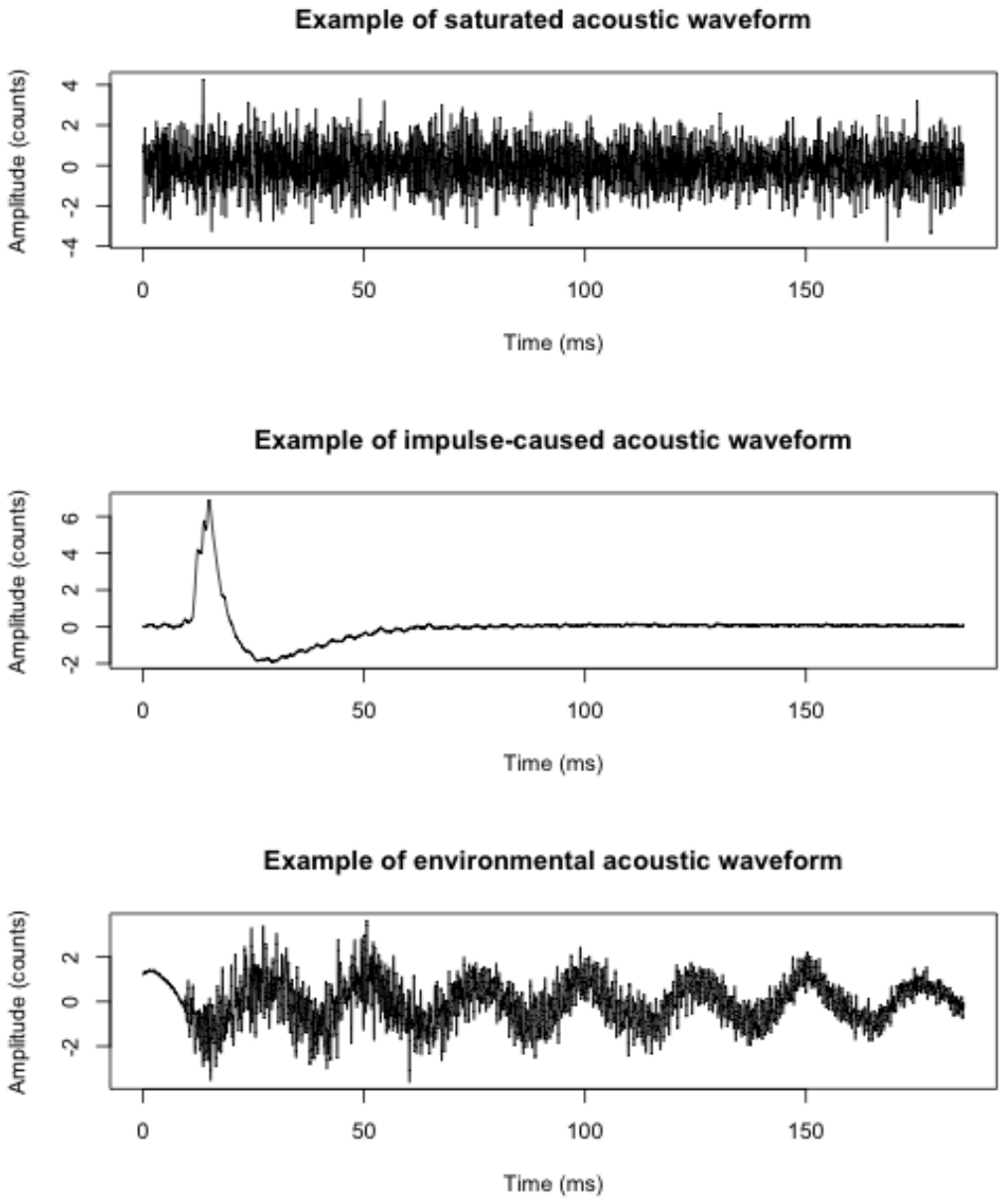

Figure 6.5: Three examples of the types of anomalous waveforms found in the operational set. The waveforms provide the examples for the following anomalies; (A) saturated waveform, (B) impulse-type waveform, (C) low-frequency waveform.

spans fewer samples than the duration of one oscillation in the low-frequency signal. This waveform is generated by environmental noise in the vicinity of the ESD and is specifically assumed to be caused by rock drilling in the vicinity of acoustic data collection. Rock drilling presents as low-frequency noise in seismic signals and is therefore assumed to present as low-frequency acoustic noise on the ESD as well. Waveforms captured that display this characteristic are assumed not to be related to rock being impacted by a sounding bar and are therefore discarded.

The understanding gained from the examination of outliers in the operational dataset and the visual characteristics of these samples are used in the data cleaning of the operational set. A total of 173 anomalous samples was identified in the operational dataset. These samples were removed before the set was used for the creation and testing of models. Boxplots showing the distribution of the cleaned dataset can be seen in Figure 6.6 on the facing page. Notice that the mean relationship for the 'safe' and 'unsafe' classes is reversed from the S_c relationship in Figure 6.4 on page 73. This new mean relationship corresponds to what is expected empirically, i.e. the pitch of 'safe' rocks is typically higher than that of 'unsafe' rocks.

Anomalous samples can be elimated during the real-time usage of the ESD by the operator who would be aware of the environmental conditions. For instance, if machinery is operated in the area the operator can discard reactions from the ESD when it is apparent that the causitive event is from the machinery noise as opposed to a strike event.

It is illuminating to relate the presence of the anomalous samples to the conditions that existed during their collection. In Subsection 5.5 on page 62 the observed conditions at the collection sites were described. Adverse conditions are expected to contribute to the amount of anomalous samples collected per data set. To illustrate this relationship, Figure 6.7 shows the ratio of 'bad' samples that are present in each subset of the operational data. As can be seen, the proportion of anomalous acoustic samples in Sets 5, 6 and 7 is more than a third of each set. Set 5 does not have conditions recorded for it. Set 6 was collected partly from the working face of the stope, as opposed to the hanging wall for the other sets. Set 7 was collected partly with a special extended sounding bar. This would indicate that the ESD is sensitive to inconsistencies in the data collection process.

Spectral descriptors are used as part of the feature set in order to provide a component in the feature set that describes the general distribution in discrete variables. The next two components of the combined feature set, namely the frequency band content features and the MFCCs, provide a more detailed breakdown of the frequency distribution of the acoustic samples for use in the models.

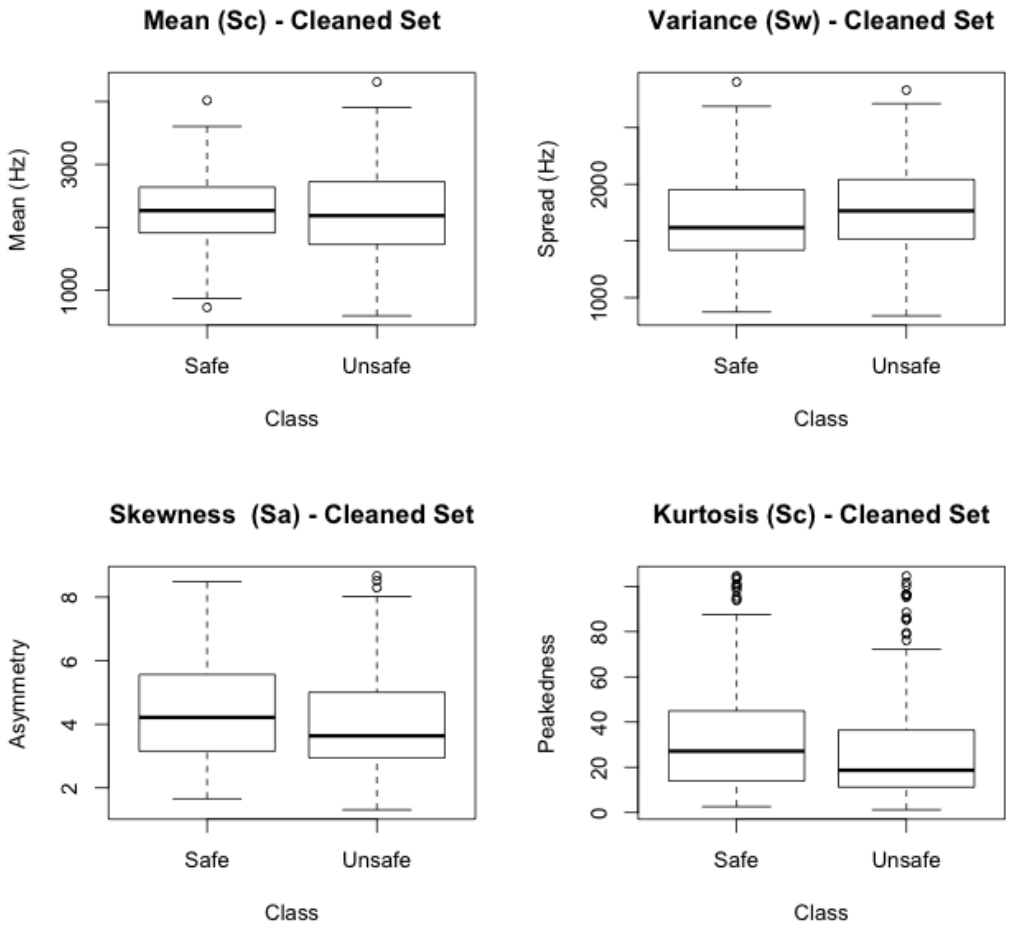

Figure 6.6: The distribution of the spectral descriptors by class values for the cleaned dataset of acoustic samples collected from Driefontein Gold Fields mine

6.4 Frequency band content

In this section the frequency band content features are discussed. These features can alternatively be called the frequency bin parameters. The variables of this feature set are the average of energy found in discretely defined frequency bands. This section derives and evaluates the feature set obtained from implementing this method on the experimental data and the operational data. The frequency band features are then added to the combined feature set for use in training and testing the models in the next chapter.

In the previous section, the spectral descriptor features were discussed. The frequency band content features differ in purpose from the spectral descriptors in that the frequency band content provide a direct representation of the content of the signal. The inclusion of these features provides the model with a break down of the energy of the signal, potentially exposing relationships that

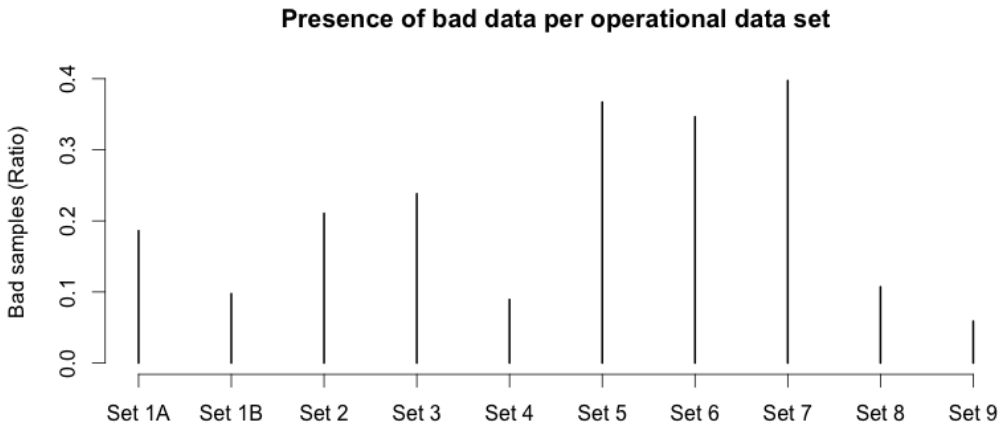

Figure 6.7: Anomalous samples as a ratio of all the samples collected for each operational set

can be used to differentiate a solid rock from a loose one.

The first step in deriving the frequency band content is to derive a frequency representation of the acoustic signal. This intermediate step before the feature transformation is used for all the methods in this chapter. The frequency transformation method is given here, but it applies to all the methods.

Each of the feature derivation methods in this chapter relies on the energy distribution of the signal represented in the frequency domain. The conventional way to derive energy in the frequency domain is through the use of periodograms. A periodogram is a standard signal transform that is used to represent an estimation of the spectral density in a time-series signal. The first step in obtaining the periodogram values for a signal is to derive the frequency domain representation of the signal. This is accomplished by applying a Discrete Fourier Transform (DFT) to the signal. For a signal, $s(n)$, with n data points, the result of the complex DFT is called $S(k)$. For the acoustic samples collected in this research, n is one of the total of $N = 4\,096$ data points per acoustic sample. The DFT is applied by the following equation over each sample where K is the length of the DFT:

$$S(k) = \sum_{n-1}^{N} (s(n)e^{-j2\pi kn/N}) \quad 1 \leq k \leq K \tag{6.4.1}$$

Each resulting $S(k)$ value of a DFT results in a complex number that represents both the phase and the amplitude of a sinusoidal component of $s(n)$. If $s(n)$ was a complex function, $S(k)$ would result in an unsymmetrical distribution. An interesting characteristic of Fourier transforms is that when applied to real signals, such as waveforms collected by the ESD, it results in

an $S(k)$ distribution that is completely symmetrical (Sorensen *et al.*, 1987). This relationship is captured as follows:

$$S(k) = S^*(N - k) \tag{6.4.2}$$

This leads to a useful corollary of the mentioned characteristic — due to both halves of the $S(k)$ distribution being identical, only half of the resulting $S(k)$ needs to be considered for practical frequency analysis. In the case of the ESD's acoustic signals of length $N = 4\,096$, the result of the DFT is a vector of $K = 4\,096$ frequency points which can be halved to $K^* = 2\,048$ points.

After $S(k)$ has been derived and halved, the absolute value of the complex Fourier transform is taken and the result is squared to obtain the periodogram estimate of the acoustic signal. This final transformation is given in the following equation:

$$P(k) = \frac{1}{N}|S(k^*)|^2 \tag{6.4.3}$$

In the case of an acoustic signal on the ESD that contains 4096 data points, this results in a periodogram estimation of half that length, i.e. 2 048 discrete frequency bins per acoustic signal. The sampling frequency of the original signal determines the frequency range represented in the periodogram estimation. The principle of the Nyquist frequency determines the effective range represented in the periodogram. The Nyquist frequency is half of the sampling frequency of the original signal. The sampling frequency of the ESD recording process is 44 100 Hz over two channels. Due to both channels being recorded on the same microphone, the two channels are duplicates of each other and therefore only one channel with an effective frequency of 22 050 Hz is used. The Nyquist frequency of half of this sampling frequency, and therefore the effective range represented in the periodogram, is 11 025 Hz.

A frequency bin the in the periodogram represents an aggregate estimation of the spectral density at a specific frequency point of the acoustic signal. As was discussed in the previous paragraph, there are 2 048 bins in the range of 11 050 Hz. Therefore each successive bin is a step size of $\frac{11\,025\text{ Hz}}{2\,048\text{ bins}} = 5,38$ Hz. This is the highest spectral resolution achievable with the sampling rate used in the data collection for this research.

If the frequency bins are used directly as input to the model of the ESD, there would then be a feature vector of 4 096 data points per sample. Very simple classification models would still be able to function with an input feature vector of this length. Any complex model, however, would require a level of processing power and duration of calculation that is not feasible on the ESD due to the limited speed of the ESD's processor. It is therefore necessary to aggregate the existing frequency bins into a smaller set for eventual use as effective feature data.

The first step in aggregating the feature bins into a meaningful and useful feature set is to define aggregate feature ranges. The averages of the peri-

odogram values in these frequency ranges are then defined as the aggregate frequency bins.

In Chapter 2, the Literature Review, the domain of psychoacoustics were discussed in Subsection 2.3.3. The role of sound perception provides a good basis for a decision on which frequency ranges should be used for the aggregation of periodogram data. The basic functionality that the ESD wants to replicate is the decision-making process of a miner when hearing the same acoustic waveform that the ESD captures. Psychoacoustics provide a guideline on human hearing, and an empirical approximation for pitch doubling that can be used as a template for determining the frequency ranges into which to aggregate the periodogram $P(k)$ values.

The frequency ranges determined by Umesh *et al.* (1999) (see Subsection 2.3.5 for discussion) were used as a template for the frequency aggregation of this feature. Table 6.1 defines the 16 frequency ranges that were used for the frequency band aggregation in this research. The variation in this table from the original data published by Umesh *et al.* is due to the frequency range limitation of the acoustic samples collected by the ESD. Figure 6.8 shows a periodogram of a collected acoustic sample from the operational set overlaid with a visual representation of the bin frequency ranges on the horizontal axis, with the corresponding average over the range indicated as the amplitude of the bin on the vertical axis.

The calculation to derive the discrete value for each bin is given by the

Table 6.1: Frequency ranges used for the Frequency Band feature bin aggregation method

Bin	Start Freq (Hz)	End Freq (Hz)
Bin 1	0	40
Bin 2	40	161
Bin 3	161	200
Bin 4	200	404
Bin 5	404	693
Bin 6	693	867
Bin 7	867	1 000
Bin 8	1 000	1 200
Bin 9	1 200	2 022
Bin 10	2 022	3 000
Bin 11	3 000	3 393
Bin 12	3 393	4 109
Bin 13	4 109	5 526
Bin 14	5 526	6 500
Bin 15	6 500	7 743
Bin 16	7 743	11 025

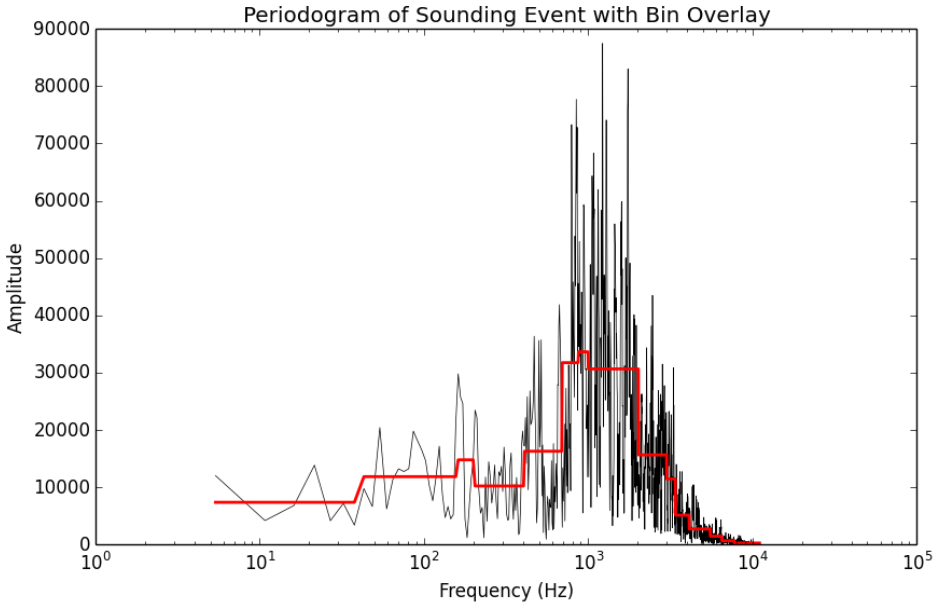

Figure 6.8: Periodogram of acoustic waveform with visual overlay of derived aggregate bin values

follow equation:

$$B(j) = \frac{1}{N_j} \sum_{k_j(min)}^{k_j(max)} P(k_j) \tag{6.4.4}$$

Each periodogram value $P(k_j)$ in Bin j is summed together, and the average value for each Bin j is found by dividing the sum by the number of periodogram samples in that range. This results in a vector of 14 discrete points. This vector is called the Frequency Band Content feature.

The Frequency Band Content feature is derived for both the experimental and operational datasets. The analysis of this operation for the two sets is given in the next two subsections.

6.4.1 Frequency band content - Experimental data

The acoustic waveforms collected from the experimental rig are processed into the Frequency Band Content feature in this subsection. The method of averaging spectral bins across predefined ranges is explained, and the results are given.

Figure 6.9 shows an error-bar plot of the 16 frequency bands that were derived from the experimental dataset. Note the difference in mean energy per band on the vertical axis. The difference between the two classes, 'safe' and 'unsafe', is particularly exaggerated in Bins 4 to 8 which correspond to

the frequency range of 400 Hz to 1 200 Hz. This observation correlates with the first boxplot of Figure 6.3 which also indicated a pronounced difference between the mean value of the two classes.

No additional data insight is gained from this set of data that has not already been made clear by the previous evaluation in the last section. The operational data is expected to expose more structural relationship in this particular analysis of the data.

6.4.2 Frequency band content - Operational data

The Frequency Band Content feature derived from the operational dataset is evaluated in this subsection. This is the most relevant application of the Frequency Band Content method as its result forms the largest vector in the combined feature set that was used in the next chapter in training models.

The method of aggregating over pre-defined bins was applied to the periodogram of the operational dataset. This resulted in a vector of 16 aggregate values for each acoustic sample in the set. Figure 6.10 shows the class-differentiated plot of the mean value of the 16 bins across all the acoustic samples in the cleaned operational set, with associated error bars indicating the 95% confidence level of each mean value.

In Figure 6.10 a much closer correlation between the two class distributions for the Frequency Band Content values can be seen compared with the correlation seen for the experimental dataset in Figure 6.9. This corresponds to the same similarity in the class distributions found in the S_c boxplot of Figure 6.6 on page 77. This implies greater difficulty for the models in the next chapter in finding a hyperplane in the features able to separate these classes, compared with the easy separation visible in the experimental dataset. Bins 8 to 12 do

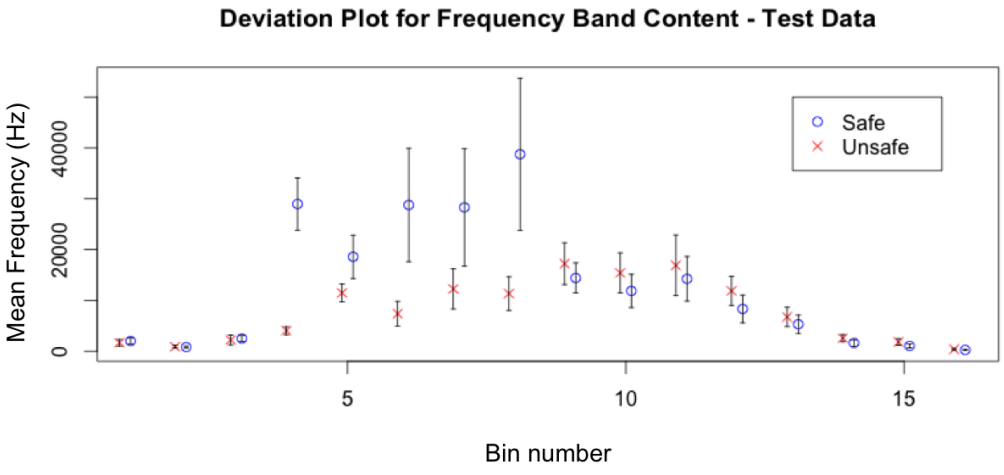

Figure 6.9: Deviation plot of frequency band content values of the test dataset

show a greater level of distinction between the classes compared with the other bins in the set. At the end of the next chapter in Section 7.4 on page 111 a test is performed to establish the relative significance of feature values in the task of classification. It is expected that Bins 8 to 12, corresponding to frequency range 1000 Hz to 4109 Hz, would be indicated as the most relevant bins in this set.

The method of aggregating specific frequency ranges in a Frequency Band Content feature helped to expose more of the spectral structure found in the acoustic samples. From visual inspection of Figure 6.10 a greater understanding of the relative spectral shapes of the features with regards to their associated class is gained. The inclusion of this feature subset in the combined feature set created in this chapter benefited the task of classification detailed in the next chapter.

6.5 Mel frequency cepstral coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are a feature approach widely used in acoustic classification algorithms to describe spectral characteristics. The use of MFCCs in the domain of acoustic classification is widespread, and is particularly dominant in the realm of speech recognition. MFCCs are also widely used in music genre classification. The particular overlap of these fields and the current problem of structural classification through the use of acoustics lies in the use of MFCCs to also identify percussion types in the aforementioned music classification. Percussion is acoustically closely related to impact events, and the use of MFCC features were found to be similarly beneficial to improved

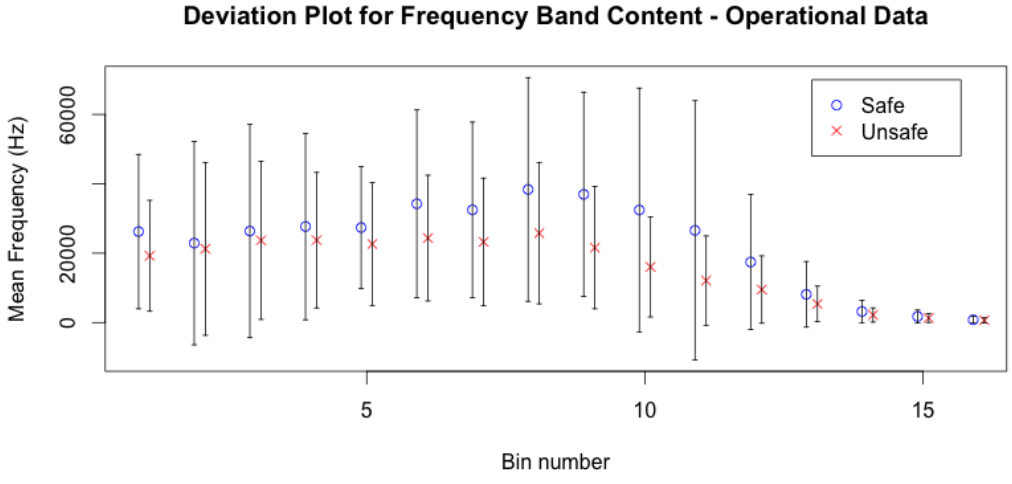

Figure 6.10: Deviation plot of frequency band content values of the operational dataset

classification in this research.

The steps in creating an MFCC are as follows:

1. Frame the signal into shorter frames.
2. For each frame calculate the periodogram values.
3. Apply the Mel filter-bank to the periodogram values and sum the energy in each filter.
4. Take the logarithm of all the resultant average energy values.
5. Take the discrete cosine transform (DCT) of the log filter-bank energies.

A few notes needs to be given regarding this process. Firstly, the signal length in one of the ESD acoustic waveforms is already irreducible. Dividing the 180 ms of a waveform length into shorter frames would be redundant for this research — it is assumed that the useful information is concentrated in the first impact envelope of the time series. Imposing frames on this waveform structure would imply that useful information is included only in the first, or first few, frames. The decision is taken to use the entire signal as one block for this MFCC derivation process.

Secondly, note that steps 2 to 4 correlate closely with the method of the previous subsection, the creation of the Frequency Band Content feature. The difference in content that these processes would deliver lies in the overlapping nature of the filter-bank used in the MFCC, in combination with the output of the DCT. This is illustrated below Figure 6.11 during the detailed discussion of the relevant step. This overlap captures more of the transient energy between frequency ranges, potentially exposing more of the critical differentiating features in the spectral envelope.

The key to understanding the approach of MFCC is to understand the Mel 'filterbank' used in the method. A short implementation description is given here for insight. Recall that the Mel scale relates perceived pitch to actual frequency. The equation for converting from frequency to Mel scale is as follows, followed by the inverse equation to convert back from Mel scale:

$$M(f) = 1125ln(1 + \frac{f}{700}) \tag{6.5.1}$$

$$M^{-1}(m) = 700(e^{\frac{m}{1125}} - 1) \tag{6.5.2}$$

These equations are used to define the Mel scale values of the frequency range of the ESD samples (Makhoul and Cosell, 1976). It will be remembered that an acoustic sample collected by the ESD has the frequency range of 0 Hz to 11 025 Hz. Applying equation 6.5.1, the maximum frequency of 11 025 Hz corresponds to a Mel value of 3 170 Mel. The usefulness of the Mel scale lies in the fact that a linear sequence in the Mel scale corresponds to a logarithmic

sequence on the frequency scale. This can be illustrated by the following example in creating a Mel filterbank with 10 filters. 10 equidistant values are first chosen between 0 and 3 170 Mel and populated in vector $m(i)$:

$$m(i) = (0, 288, 576, 865, 1\,153, 1\,441, 1\,729, 2\,018, 2\,306, 2\,594, 2\,882, 3\,171)$$

Applying equation 6.5.2 to each point in $m(i)$, the following values for frequency vector $f(i)$ are calculated:

$$f(i) = (0, 204, 469, 810, 1\,251, 1\,820, 2\,556, 3\,507, 4\,736, 6\,324, 8\,375, 11\,025)$$

Note that the last value is 11 025 Hz and that $f(i)$ spans the entire ESD frequency range in a logarithmic distribution. This creates the basis for the Mel filterbank. To construct the filterbank, a triangular function is applied to each frequency point $f(m)$ in $f(i)$ according to the following equation:

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq (m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases} \tag{6.5.3}$$

To assist in the understanding of this equation and the concept of the triangular filterbank, a plot is created that shows all 10 filterbank overlaid on each other. This is shown in Figure 6.11. As can be seen, each filterbank vector is non-zero for a certain part of the spectrum. To calculate filterbank energies each filterbank is successively multiplied by the periodogram values of the acoustic samples, and then the coefficients are added up. Once this has been performed, a total of 10 values remain which give an indication of the amount of energy per filterbank. The log of each of the filterbank energy values is taken to normalise the values to a logarithmic distribution.

Up to this step in the MFCC derivation process, it follows a similar structure to that of the previous subsection, the Frequency Band Content feature derivation. As pointed out, one difference between the methods is the dynamic overlapping filterbank that has just been defined. The second main difference between the methods is the next step of the MFCC creation process. An additional transform is now done on the log-energy data. The DCT is calculated for each log-energy value created by the filter bank. The DCT has very good energy compaction, which means that the it concentrates the most important information in the lowest bins of the resultant coefficient set, with each higher bin holding less than the bin before it. This is why DCT is so often used in compression. The advantage of this in classification is that it is only necessary to compare the first few resultant coefficients of the DCT to get a measure of how similar two acoustic waveforms are.

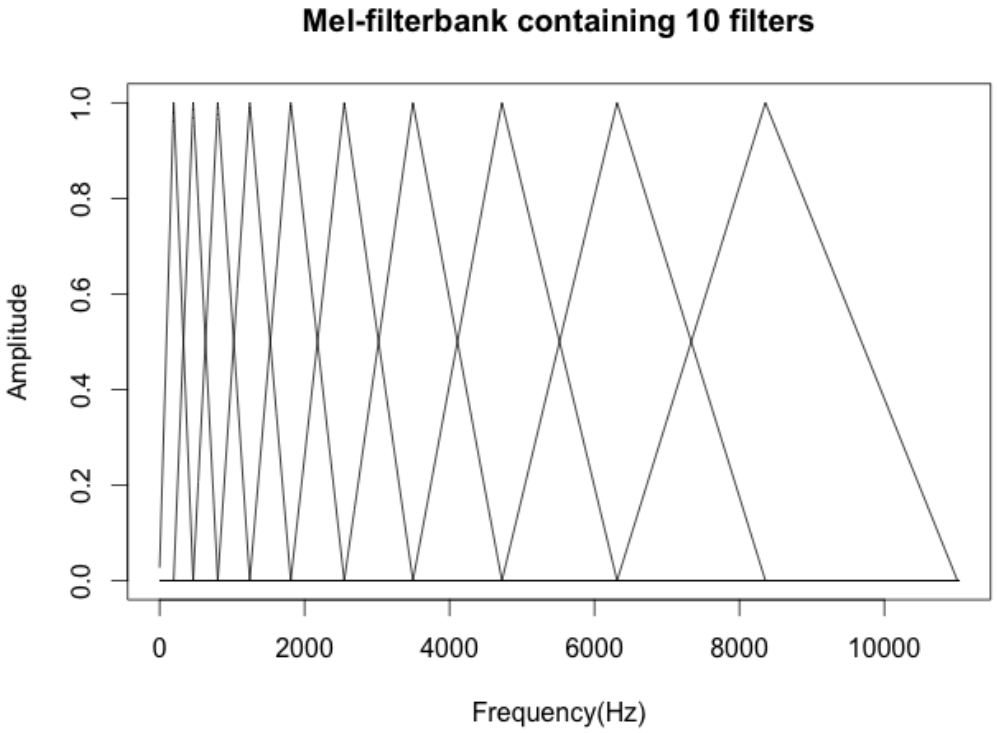

Figure 6.11: A Mel-filterbank containing 10 filters — This filterbank starts at 0 Hz and ends at 11 025 Hz

The version of the DCT equation chosen for use here is the most common variant of the DCT, called the type-II DCT. Its equation is given as follows:

$$DCT_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2}\right)k\right] \tag{6.5.4}$$

The final result of this is called the MFCC values of the acoustic waveforms. An example of what a resultant distribution looks like is shown in Figure 6.12. It is not as easy to interpret the distribution shape of the final transformation visually as as it was with the spectral distribution, due to the concentration of information in the first few coefficients. A property of the DCT coefficients, and therefore of the MFCC, is that the zeroth bin contains a crude approximation of the amount of energy in the entire signal. This bin is optional for a feature set. In human language translation (a large application domain for MFCC) the zeroth bin is typically discarded. The choice was made to retain this bin for this research, as it has contributed to improved classification results in percussion classification by Gillet and Richard (2004), as it has proved to be a very significant feature in this project as will be shown in Section 7.4 on page 111.

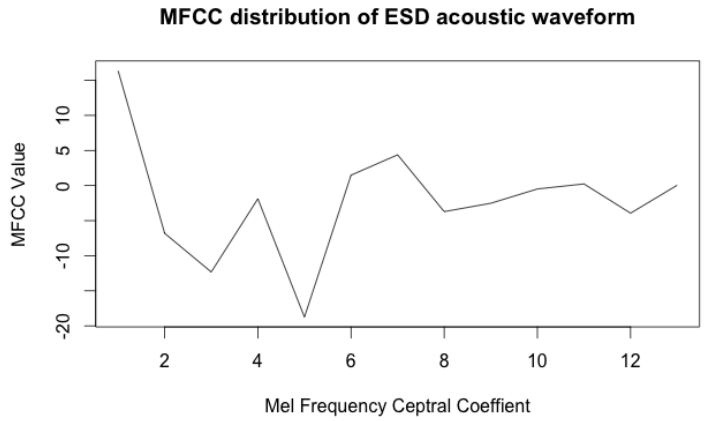

Figure 6.12: Plot showing the MFCC distribution of an acoustic waveform from the operational data set

This process is applied to both the experimental and the operational data sets. A distribution plot of the MFCC as seen in Figure 6.12 does not provide insight into the data as it did in the previous sections. MFCC values are difficult to relate back to spectral features due to the additional transformation which effectively calculates the spectrum of the spectrum. However, visualisation of the MFCC feature is still useful in order to relate the difference in structure between the classes and dataset.

The relative structures of MFCC distributions between samples will be used by the models for comparison and classification. A simple example of accomplishing this is by deriving an Euclidean distance between the MFCC points of two acoustic samples to provide a measure of the similarity of the two samples.

Figure 6.13 on the following page provides an insight into the relation of the mean of the first ten MFCC values differentiated between the 'safe' and 'unsafe' classes for both the experimental and operational data sets. This shows that there are differences in structure within the distribution of the MFCC values of the two separate classes. The distinction is evidently more pronounced in the test rig data set than the operational dataset — a result expected due to similar differences seen in the previously derived features.

The MFCC values are the last feature to be derived for the purpose of this research. In the next section, the three derived feature groups are combined into the X_{input} feature matrix.

6.6 Combined feature set

The purpose of this entire chapter is to describe the feature set that will be used to train and test the models in the next chapter. The previous sections

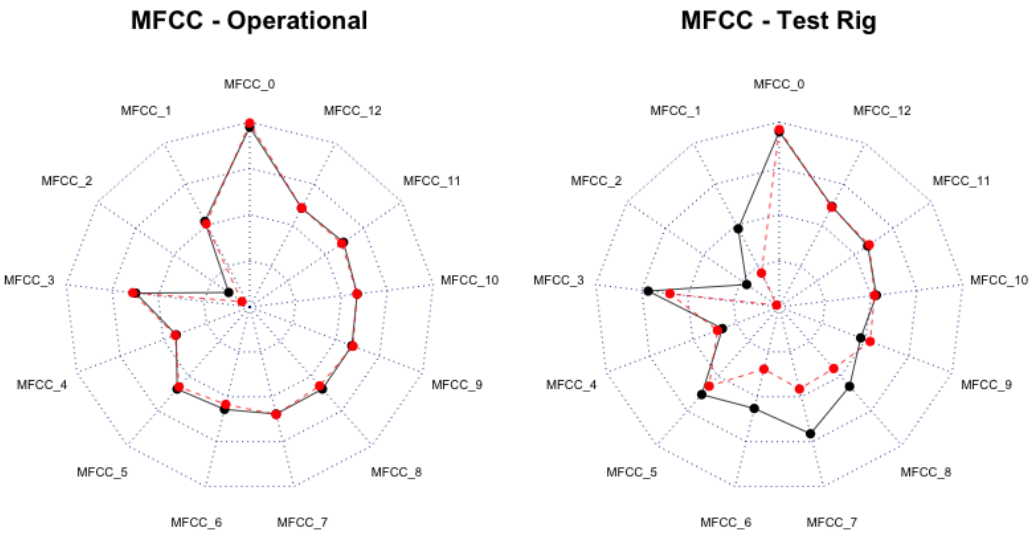

Figure 6.13: Radar plot showing the mean of the first 12 MFCC values differentiated by class and dataset. The black line indicates the 'safe' state, and the red line indicates the 'unsafe' state

of this chapter detailed the three feature approaches that are derived from the ESD acoustic waveform structure. Additional analysis and preparation of the combined feature set of the operational data will result in a functional basis for the next chapter. The purpose of this section is to describe the combined feature set X_{input} of the operational data, normalise the features and test for excessive correlation between features.

The combined feature set is defined as a matrix named X_{input}. This matrix is defined as a row for each acoustic sample's features, and the columns are the types of feature. Table 6.2 on the next page show the features and their corresponding labels in the combined feature set. It also shows the aggregate statistics for each feature, i.e. mean and first and third quantile values. Note the existing ranges and distribution for each feature. These values are the direct outputs from the relevant equations and transformations necessary to construct the feature, and therefore the individual feature groups do not adhere to the same universal range. Most of the models that will be investigated in the next chapter will not function well, or at all, with the observed ranges in the feature data. Even the models that can accept a broad feature range will be negatively affected due to the difference in the order of magnitude of the ranges observable between sub-feature groups.

Table 6.2: Complete operational feature set with mean and deviance parameters

X_{input}	Feature Group	Label	Mean	1st Quartile	3rd Quartile
X_1	Descriptor	Sc	1 797.0	592.2	2 261.9
X_2		Sw	1 455.0	839.6	1 751.5
X_3		Sa	3.0	1.3	4.2
X_4		Sf	12.1	1.2	29.9
X_5	Band	mel_bin1	9 793.0	687.0	22 061.0
X_6		mel_bin2	7 373.3	670.2	21 919.4
X_7		mel_bin3	9 403.1	392.8	24 781.1
X_8		mel_bin4	11 755.8	660.6	25 343.1
X_9		mel_bin5	13 266.0	1 104.0	24 546.0
X_{10}		mel_bin6	14 622.0	1 404.0	28 335.0
X_{11}		mel_bin7	13 422.4	978.4	26 972.7
X_{12}		mel_bin8	14 412.0	1 430.0	30 845.0
X_{13}		mel_bin9	11 715.0	1 842.0	27 772.0
X_{14}		mel_bin10	8 152.1	990.5	22 646.4
X_{15}		mel_bin11	4 922.8	617.1	17 930.0
X_{16}		mel_bin12	3 591.9	340.4	12 716.0
X_{17}		mel_bin13	1 845.9	243.9	6 490.1
X_{18}		mel_bin14	782.2	65.7	2 605.5
X_{19}		mel_bin15	442.3	56.9	1 493.4
X_{20}		mel_bin16	174.4	33.4	722.2
X_{21}	MFCC	MFCC_0	15.9	13.7	17.1
X_{22}		MFCC_1	-11.5	-25.3	-5.2
X_{23}		MFCC_2	-31.0	-47.4	-25.0
X_{24}		MFCC_3	-3.0	-19.7	0.2
X_{25}		MFCC_4	-13.2	-26.4	-9.5
X_{26}		MFCC_5	-5.6	-20.8	-1.9
X_{27}		MFCC_6	-7.0	-20.0	-3.2
X_{28}		MFCC_7	-4.4	-15.0	-1.3
X_{29}		MFCC_8	-4.9	-20.5	-2.0
X_{30}		MFCC_9	-4.0	-17.7	-1.5
X_{31}		MFCC_10	-4.1	-13.3	-1.9
X_{32}		MFCC_11	-2.8	-11.2	-0.6
X_{33}		MFCC_12	-3.2	-11.0	-1.0

6.6.1 Data scaling

In order to make the combined feature set usable by most models, the ranges of each feature need to be bound within a specific range. In Chapter 3, Methodology, the terms 'standardisation' and 'normalisation' were defined, on page 41. These relate to the data preparation process called data scaling. The purpose of data scaling is to transform the feature set into a defined value range acceptable for use in models. This imposed range is the difference between standardisation and normalisation. In the case of normalisation, $X_{norm} \in [0, 1]$. For standardisation, $X_{stand} \in [-1, 1]$. The wide variety of models available differ as to which of these two data scaling methods is necessary or optimal. In this research, both of these data scaling techniques were applied to the feature X_{input}, and all models were subsequently trained and tested with both resultant sets. This agnostic approach to the two methods of data scaling then reflect the optimal combination of feature scaling and model in the final results. It also ensures equal representation of each feature during training.

The normalised data are derived through the min-max equation:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{6.6.1}$$

The equation to derive the standardised feature is easy to apply once the mean μ and the variance σ of the feature is known:

$$X_{stand} = \frac{X - \mu}{\sigma} \tag{6.6.2}$$

The equation for X_{stand} results in a distribution centred around 0 with a standard deviation of 1. The values of μ and σ determined by examining the training set of the operational data is applied to the testing data, and subsequently X_{stand} is codified explicitly on the final implementation of the ESD. This is especially useful in models that measure distance between samples, and is expected to lead to improved performance in the results with these types of models.

6.6.2 Feature correlation

Three methods of constructing features were implemented in this chapter, independently of each other. This resulted in the final combined feature set consisting of the values derived through all three methods. This independence in approach does not, however, result in three sets of values that are statistically independent of each other. Some correlation between certain features is expected, considering the overlapping goals of the approaches in reducing the dimensionality of the acoustic waveforms. This subsection aims to explore the correlation between variables.

Ideally, a scatterplot matrix comparing all the feature variables against each other is constructed to indicate visually the measure of similarity between

features. In smaller feature sets this is the typical approach, and it provides intuitive understanding of the potential correlation between variables. In this research there are 33 feature variables, as can be seen in Table 6.2. This large number of features would result in scatterplots shown in an array of 33 rows and 33 columns - this is unwieldy for the sake of illustration and intuitive understanding.

An alternative approach is to derive the statistical correlation coefficients of each feature for comparison against each other. This will result in a covariance indicator that shows how two features change together. The covariance indicator is intended to show whether a linear relationship can be found between any two features. In statistics, a standard approach is to use the Pearson's correlation coefficient matrix to determine this relationship (Lee Rodgers and Nicewander, 1988).

The equation for Pearson's correlation coefficient r_{xy} between any two variable distributions, or commonly just called the correlation coefficient, is as follows:

$$r_{xy} = \frac{\sum x_i y_i - n\mu_x\mu_y}{(n-1)\,\sigma_x\sigma_y} \tag{6.6.3}$$

where μ_x and μ_y are the means of features X and Y, σ_x and σ_y are the standard deviations of X and Y, and n is the amount of samples in the individual variable distributions.

Deriving the correlation coefficient for the feature set results in a matrix of values with dimensions of 33 rows by 33 columns. Each of the points on the matrix corresponds to the correlation of the two features identifying the corresponding row and column. Applying the colour coding to the values in the range of [-1, 1] helps in the quick identification of values by visual inspection. The resulting plot is given in Figure 6.14.

The diagonal line of Figure 6.14 shows the full correlation of each feature with itself, so is discarded for this analysis. The first strong correlations that can be seen are those of S_c with S_w, and also S_a with S_f. A higher mean in the distribution corresponds to a higher variance, which indicates that the spread of spectrum of the acoustic signal becomes less detailed in the higher frequencies. The peakedness also correlates strongly with the skewness of the signal; these are properties that are influenced by the higher definition of the spectrum in the lower frequencies. This observation relates to discussion around the perception of pitch, where lower frequencies are more iteratively distinguishable than higher frequencies.

A strong correlation can be seen between each neighbouring group of frequency bins. The existence of energy in the frequency ranges is therefore not strongly differentiated by individual bins. This indicates that there are transient values between the bins, which will assist in providing additional detail regarding the frequency distribution as described by these bins.

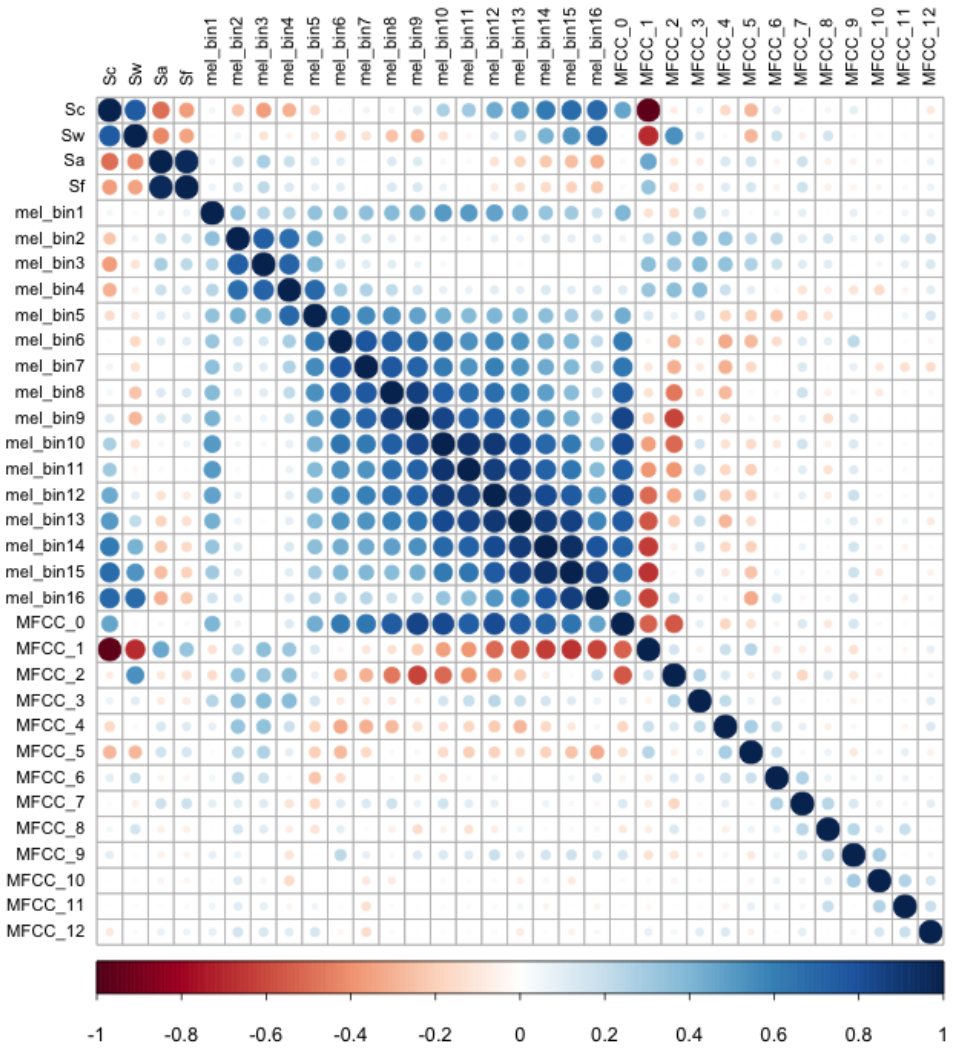

Figure 6.14: Correlation coefficients for the feature set depicted in a colour chart

The strongest correlation can be seen between the mean of the signal, S_c, with the first-order MFCC value, $MFCC_1$. This correlation approaches negative unity, which indicates that the higher the mean of the frequency of the signal, the lower the value of the first MFCC. This is a surprising correlation, and provides additional insight into the use of the MFCC approach to describe the higher-order characteristics of signals.

The correlation matrix shows that no individual feature is redundant to the feature set. Each feature describes a different aspect of the acoustic signal. This justifies the use of this feature set for the training of models, and the additional information provided by each feature variable will ideally increase the accuracy of the model's construction in the next chapter.

6.7 Conclusion

In this chapter the acoustic waveforms collected from the experimental rig and from operational environments were transformed into a descriptive feature set. This feature set describes important and distinguishable characteristics of the acoustic waveforms in a vector of 33 values. This compares very favourably with the original acoustic waveform length of 4 096 data points: the feature set effectively reduces the dimensionality of the signal. This reduced dimensionality aids in the construction, training and testing of potential classification models.

In addition, the feature set values were standardised into consistent data ranges. This is essential for the operation of most models.

The successful preparation of a feature set is the final step before the training and testing of models. The intended use of the feature set will be shown in the next chapter.

Chapter 7

Classification Models Evaluated

7.1 Introduction

This chapter describes the choice of the appropriate classification model for use on the ESD. To paraphrase the No Free Lunch Theorem, there is no single 'best' model, and each problem needs a tailored classification system to describe it best (Wolpert and Macready, 1997). Therefore, the appropriate model for this problem needs to be found from the available pool of classification models that exist in practice. This chapter aims to present a classification system that can be justified for use on the ESD.

The approach used in order to construct and evaluate models against each other is described first. The discussion focuses on the areas of feature resampling, the training methodology, and the performance measure used for scoring models.

The second part of the chapter presents the results from the model evaluations. It describes the results of the testing of classification models grouped as a common modality, after which the best result from each group is compared with each other. Top performing approaches are then identified and further optimised until a suitable classification system is found.

Finally, the chapter describes the implementation details of the chosen classification model on the ESD platform.

7.2 Model testing approach

The evaluation of different models against each other needs to be done in a consistent manner for the resulting choice of classification model implemented on the ESD to be justified. The competitive testing of models is complicated by the wide variety and different orders of implementation complexity of algorithms that can perform classification. This necessitates the specification of a formal testing environment that will ideally result in the identification of

an algorithm or set of algorithms that can be said to perform 'best' at ESD acoustic classification.

The specification for the testing environment includes the separation of the data that will be used to train and test the models. It also includes the performance measure that is used to assess the model. The goal of the testing environment is to be able to test the models quickly and consistently against a fair representation of the ESD acoustic data. The outcome of the testing environment is the knowledge of which models to consider for further optimisation.

The result of the testing environment also provides a measure of how 'learnable', or suitable for classification, the problem is. Thus far in this dissertation the holding assumption was that the ESD functionality can be obtained through classification. The results of the testing environment show whether the acoustic data are indeed a measure for determining rock structural stability. If a variety of different learning algorithms perform poorly, it is an indication that there is a lack of necessary differentiating structure to be learned.

7.2.1 Feature resampling

The main hidden obstacle in classification training is the problem of overtraining. To illustrate how damaging overtraining of a model can be to its operational performance, consider the following example. A very simple model can be constructed from feature data that purely memorise the training data. Effectively, the model is a matching template to the training data that knows the predicted class value of each given sample. This model would have perfect accuracy in subsequently predicting the training set, but would inevitably perform very badly on any unseen data for which it does not have an exact match. This model is said to be unable to generalise. It is highly important to realise generalisation in this model testing environment.

k-Fold cross-validation is the standard rigorous technique in training models. Each time the algorithm is run, it will be trained on 90% of the data and tested on 10%, and each run of the algorithm will change which 10% of the data the algorithm is tested on. It is the technique chosen for the use in the specification for the testing environment to provide an unbiased assessment of the accuracy of the model for the ESD. This measure of accuracy was used to determine the optimal tuning parameters for each model. However, as will be explained in Subsection 7.2.3, the models that are competitively evaluated against each other should still be measured against an independent testing dataset that the models were not exposed to during training. Therefore a separate testing set is defined, and it is kept separate from the training set. The final measurement of performance is done on the testing set, and this measure will be used to determine statistically significant differences between the performance of the models.

For the test environment, a ratio of 80:20 was chosen for the subsetting of the feature data into a training set and a testing set. This effectively means that for the cleaned operational set of 537 samples, the resulting training subset has 403 samples, and the testing subset 134 samples. The sets were chosen in such a way that the class distribution present in the operational dataset was replicated in both the testing set and the training set.

7.2.2 Performance measure

In order to compare different classification models with each other, a performance measure indicating their fitness needs to be defined. Various well-defined classification performance measures exist, so the task was to choose one relevant to the functionality of the ESD.

Classification models do not typically generate a discrete prediction, but rather the probability between 0 and 1, which is then interpreted by some threshold as a discrete prediction. For instance, a model could return a value of 0.71, which would be rounded up and interpreted as the class 1. In the particular case of the ESD classification requirements, there are two predicted classes, 'safe' and 'unsafe'. The class value of 1, alternatively called the 'event', is typically associated with the predicted class that is of most interest to the problem. The ESD is designed to mitigate rock-fall risk, and rocks determined to be in loose structural cohesion are directly responsible for rock falls. In the case of the ESD, therefore, it was decided that the class of 'unsafe' is the event of interest for detection. Hence, 0 = 'safe' and 1 = 'unsafe'.

A useful way to present the results of a model's prediction is in the form of a confusion matrix. A confusion matrix is a cross-tabulation of the observed and predicted classes for the data. The confusion matrix for this research is shown in Table 7.1. The first value is called the True Positive (TP), and it is the number of samples that were correctly predicted to be an event. True Negative (TN), correspondingly, relates to the correct prediction of a non-event. A False Positive (FP) is the case where an event is predicted, but in fact there is no observed event. An example of an FP is when the ESD predicts an 'unsafe' state when the rock is in fact secure. A False Negative (FN) occurs when the prediction is for a non-event, but in reality there is an event. An FN example is when the ESD presents a 'safe' prediction when the rock mass is dangerously loose.

Table 7.1: Confusion matrix template indicating the number of true positives (TP), false positives FP, true negatives (TN) and false negatives (FN)

	Real 'unsafe'	Real 'safe'
Predicted 'unsafe'	TP	FP
Predicted 'safe'	FN	TN

Two additional statistics relevant to the performance measure for this research can be derived from Table 7.1. The sensitivity of a model is the rate at which the event of interest, i.e. 'unsafe', is predicted correctly for all samples having the event, or

$$Sensitivity = \frac{TP}{TP + FN} \tag{7.2.1}$$

The sensitivity of the prediction measures the accuracy in the event population. The converse statistic is the specificity of the model. The specificity is defined as the rate that non-event samples, i.e. 'safe', are correctly predicted, or

$$Specificity = \frac{TN}{FP + TN} \tag{7.2.2}$$

Sensitivity and specificity tend to be inversely related to each other in most practical model outcomes. An ideal model with perfect prediction, and hence perfect TP and TN values, would have a value of 1 for both these statistics, but that rarely happens and did not happen in this research.

Recall that model outputs are not discrete numbers, 1 and 0, but rather lie on a continuous scale between these two values that indicate probability. The assumption does not hold that rounding the output number to either 1 or 0 at the 0.5 threshold is ideal - in fact, the choice of threshold directly affects the sensitivity and specificity of the model. Intuitively, this can be understood through the following example. If the chosen threshold is virtually zero, then the model outcome would predict all samples to be 'unsafe' and no false negatives would occur, and hence the the sensitivity of the model would be perfect. In this same example, no true negative values would be predicted, and the specificity of the model would calculate as zero. This relationship between the variable threshold and the values of specificity and sensitivity can be visually presented in a plot called the Receiver Operating Characteristic curve (Altman and Bland, 1994), typically called by its abbreviation, the ROC curve.

The ROC curve will be used as a quantitative assessment of the models in this chapter. Figure 7.1 shows an example of a ROC curve. A perfect model that predicts completely accurately would have 100% sensitivity and specificity. Visually, the ROC curve would be a single step between (0, 0) and (0, 1), and then remain constant between (0, 1) and (1, 1). The area under the ROC curve for a perfect model would therefore be one. A completely ineffective model results in a ROC curve that closely follows the 45° diagonal line and has an area under the ROC curve of 0.5. The useful characteristic of this is that it is possible to compare different models visually if their ROC curves are superimposed on the same plot.

Referring to Table 7.1 again, attention should be paid to the FN field. In this research, a false negative would occur if the ESD predicts a rock mass to be

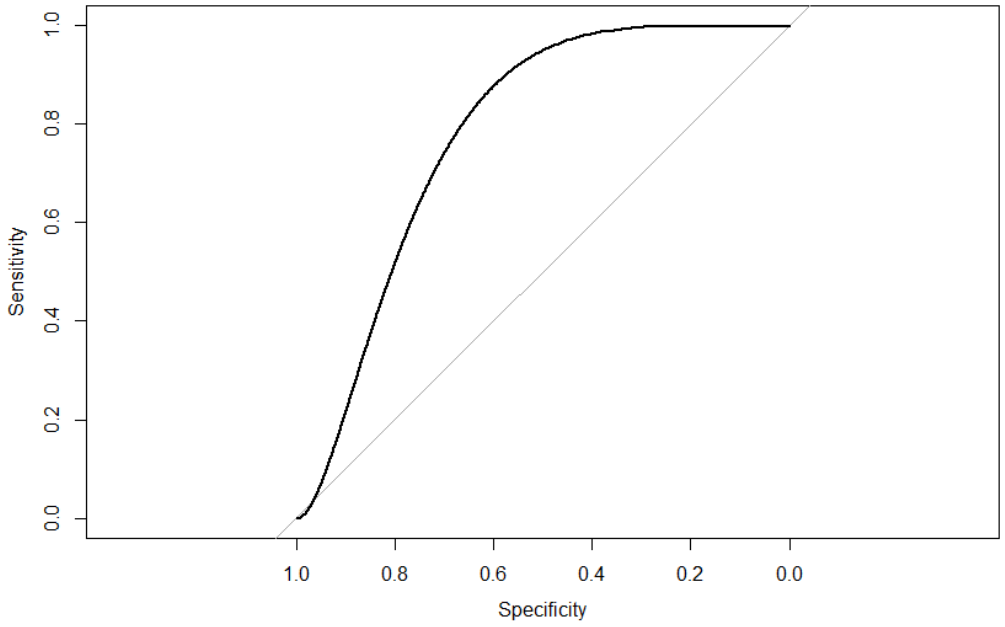

Figure 7.1: Smoothed ROC curve of example model showing the relation between specificity and sensitivity with diagonal reference line

structurally cohesive when it is actually loose. This is a category of prediction with a potentially dangerous result. Arguably, a misleading prediction is worse than no prediction at all due to the possibility of convincing the miner that an existing rock-fall risk can be ignored. The statistic that relates to this FN value is the sensitivity of the model. In the choice and optimisation of the model, its sensitivity will be the most important consideration. As an extension of this decision, the area under the ROC curve will be used as the main performance measure when evaluating models against each other.

7.2.3 Model testing environment

The model testing environment was implemented in the R programming language. This language was chosen after evaluating the different implementation options discussed in Section 3.3.7 on page 46. R is an open source software programming language that is widely used in the science of classification. The main strength of R is in its reliance on user-submitted modules to extend the language for specific tasks. In particular, virtually all state-of-the-art models currently in use are very well represented in the package repository for R.

In fact, the wide variety of models submitted by various authors presents a problem for implementation. R does not impose a function structure for models, and therefore each model package has different implementation parameters and usage. This complexity is effectively masked by the caret package

developed by Dr Khun who is the Director of Non-clinical Statistics at Pfizer Global R&D (Kuhn, 2008). The caret package (short for **C**lassification **A**nd **RE**gression **T**raining) creates a unified interface for modelling and prediction in R, and interfaces to 147 regression and classification models.

The additional benefit of caret is the helper functions it implements to assist in the streamlining of model tuning using resampling. These include functions to split the dataset into testing and training datasets that each still maintain the class distribution of the whole. Furthermore, the use of k-fold cross-validation during the training of models can be specified. A useful function provided by caret is the search for optimal model-tuning variables. Some models have specific tuning variables that can be set prior to the training of the model that make the particular model potentially more suitable for describing the problem. The specific tuning variables applicable to each type of model are described during the presentation of the model training results in the following section.

The algorithm to train and select models is as follows:

```
Split feature set into testing and training set;
Define sets of model tuning variables to evaluate;
for each tuning variable do
    for each cross-validation k-fold of training set do
        Hold-out specific fold of training data;
        Fit the model model on the remainder;
        Predict the hold-out samples;
    end
    Calculate the average performance across hold-out predictions;
end
Determine the optimal tuning variables based on highest cross-validated
accuracy;
Apply model to testing set to predict outcomes;
Construct ROC curve and compare against other models;
```

This algorithm describes the process to train each model, and provides a measurement of its ability to accurately predict outcomes based on data the model was not exposed to during its training. This provides a measure of the real-world applicability of the constructed model on the ESD in operational environments.

7.3 Model evaluation

All the models in this section are competitively evaluated on performance, and the most promising models based on their predictive result are chosen to be optimised in the next section. This section's results are grouped by model

similarity, in order to improve readability as well as to relate similarity in results to specific modelling techniques.

The grouping of models in this section is based on the fundamental approach to classification which provides a measure of similarity between them. This approach is one of several grouping techniques, although there is no authoritative model grouping in the literature. This is due to the fact that there are models that can easily fit into several categories. As such, there are many variations in the way that algorithms are grouped, depending on the source compiling the similarities. The approach followed in this dissertation is intended to group a basic classification model and an advanced classification model that are fundamentally related to each other in an intuitive sense.

A brief overview of each model is given during the relevant discussion of the setup of the particular model. Table 7.2 provides a guide to the tuning parameters evaluated during the training of each model. A more thorough discussion of the theory of each classification model is given in in Section 2.4 of the Literature review chapter. In the discussion of each classification model group, the implementation details that are pertinent to the performance of the model on the ESD are presented.

7.3.1 Instance-based models

Instance-based models aim to predict new samples by comparing them directly with a memory of training samples. A distance function between the new sample and the stored samples provides the probability measure of the prediction made by the classification model. In this subsection two instance-based methods are implemented and evaluated. The first method is the relatively simple kNN method, and the second method is the more advanced SOM classification model. In Subsection 2.4.1 the principles of these methods are discussed.

The kNN model develops a 'memory' of training samples. The only measure of training needed for a model is finding a consistent representation of the samples. Each point on the kNN map is a single sample with n dimensions, where n is the number of individual feature points. The kNN model has one

Table 7.2: Overview of the classification model tuning parameters evaluated during training

Grouping	Model	Model Parameters
Instance-based	k-Nearest Neighbours	$k \in [2\text{–}21]$
	Self Organizing Map	$xdim, ydim, xweight \in [1\text{–}10]$
Decision Trees	CART	$C_p \in [0\text{–}0.1]$
	Random Forest	$trees \in [250\text{–}2000], interaction.depth \in [1\text{–}5]$
Function	Logistic Regression	$\lambda \in [10^{-4}\text{–}10^{-1}]$
	Backprop-NN	hidden layer nodes $\in [5\text{–}20]$
Kernel	SVM - RBF	$\sigma \in [0.01\text{–}0.03], C \in [0.25\text{–}64]$
	SVM - Polynomial	$d \in [1\text{–}5], C \in [0.25, 0.5, 1.0]$

tuning variable: k is an integer indicating how many nearest neighbours should be referenced. The optimal value of the tuning parameter k is highly problem dependent. The rule of thumb for k is that a higher value suppresses the effects of noise in the training dataset but makes the classification boundaries more distinct. The kNN implementation used in this research to evaluate its suitability on the ESD data is the R package designed by Ripley and Venables (2013).

Fitting the kNN model with various tuning variables results in Table A.1 on page 127. From this table, a trained model with an optimal k-value of 5 is chosen. The range of k-values tested was from 2 to 21. The resampling results, i.e. the predicted results on each of the k-folds during cross-validation, are shown in the top line of Figure 7.2.

SOM classification models can also be said to be memory based, though the samples are not directly recorded as in the case of kNN. Each sample is approximated in a low-dimension topological map that can be visualised. The map consists of nodes that are equivalent to the neurons used in neural networks (discussed in Subsection 7.3.3), and the weights between the nodes provide the measure of similarity of samples. Training of a SOM classification model involves constructing the map by fitting the layout of individual neurons to approximate the spatial layout of the sample space, thus loosing the higher dimensionality by replicating the same structure in a simpler representation.

The SOM implementation used for comparison for this research was devel-

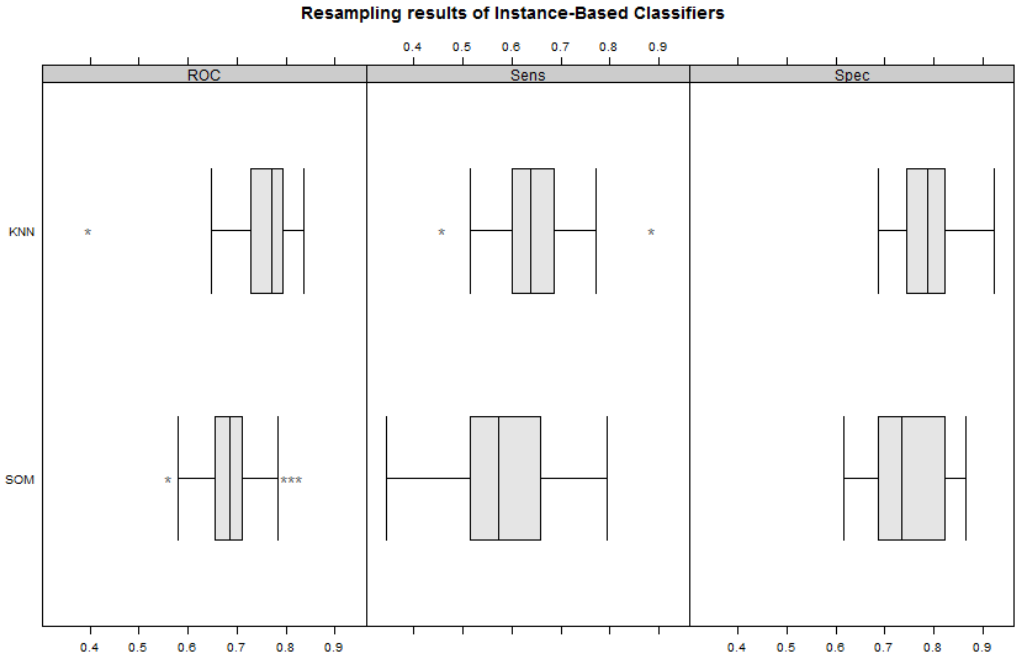

Figure 7.2: Resampling results of instance-based models showing the mean and deviance of the ROC area under curve, sensitivity and specificity

oped by Wehrens and Buydens (2007) in R. The SOM model has the most tuning variables of all the models under consideration. Four tuning parameters are present; the width and height of the neuron field, $xdim$ and $ydim$, the weight of the class values during supervised training, $xweight$, and a flag indicating whether the map can 'wrap' around on itself.

The final contender model was optimally tuned with $xdim = 3$, $ydim = 4$, $xweight = 0.5$ and the edges of the map joined. The results of varying the tuning parameters are presented in Table A.2 on page 128. The resampling results obtained from predicting on the cross-fold hold-out set during model training is shown in the second line of Figure 7.2 on the preceding page.

A comparative analysis is done on the models of this section to give an indication of the performance for an instance-based classification approach. This comparison is provided at the end of this section in Figure 7.7 on page 109, where the ROC curves of all the classification models are shown together. The kNN model has an area under curve (AUC) value of 0.7037, and the SOM model a value of 0.6819. Based on the observed distribution of the ROC curve for the kNN model, it can be seen that the kNN model has a stronger tendency towards better sensitivity performance than the SOM model. The fact that kNN follows a recall approach to stored ESD samples as opposed to the SOM's clustering approach might be reason for this difference.

7.3.2 Decision tree models

Decision tree models provide a rule-based insight into how a prediction is derived from a model. This subsection evaluates the performance of a single tree model and a Random Forest that consists of many trees. The theory behind a tree-based classification approach is given in Subsection 2.4.2.

Figure 7.3 shows the CART method in action on the ESD feature set. Note that the first split attempted the biggest separation in class values based on a value of the *MFCC_0* feature. Subsequently, a split was identified and performed on the S_a feature with the same intention. The number of splits in a decision tree determine the tree depth. The model stops training once a preset depth is reached, or a cost-complexity parameter C_p is used to monitor performance and stop training when information starts to be lost in the process. In fact, trees that are constructed to have a maximum depth are notorious for overfitting the training data. A more generalisable tree model is one that is a pruned version of the initial tree.

The CART algorithm attempts to find the best tree for a particular value of C_p. As with other cost-complexity parameters, smaller penalty values for C_p tend to produce more complex models, which in this case results in larger trees. To find the best tree, the model was evaluated over a group of C_p values.

The implementation used to develop the competitive model for this subsection for the R environment was developed by Therneau *et al.* (2014). The

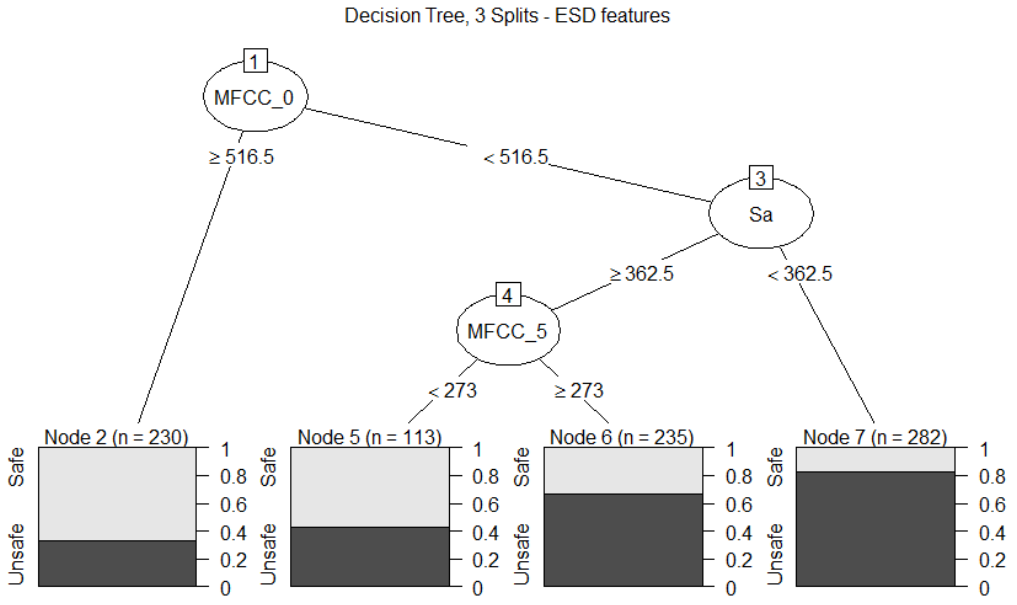

Figure 7.3: Example of the application of a decision tree on the ESD features showing the first three splits of the data

only tuning variable for this model is the complexity parameter C_p, which was initiated at zero and bought upwards by incremental fraction steps of $^1/_{300}$.

The CART model was fitted to the training data, and the optimal model was obtained. The optimal model had a C_p value of zero. This is not ideal, since it could imply a case of overtraining. Table A.3 on page 129 shows the performance obtained by training the model across the possible tuning parameters. The choice was made to discard $C_p = 0$ as an option during fitting to avoid the possibility of overfitting. The subsequent optimal value found was $C_p = 0.00777$. This resulted in the resampling statistics shown in Figure 7.4. The optimised model was applied to the testing set, and the final ROC curve was obtained with an AUC value of 0.7451. The ROC curve is included in Figure 7.7 at the end of the section.

The Random Forest model was also evaluated in this model category. The implementation for this is the R implementation by Liaw and Wiener (2002). Each tree in the Forest is assigned a subselection of feature points to evaluate, hence the randomness in the name of the model. There are two tuning variables for this implementation: the number of trees to include in one Random Forest and the interaction depth allowed for each tree, denoted by *interaction.depth*. The number of trees evaluated was from [250–2 000], and the interaction depth is tested is for 1, 3 and 5 levels.

The Decision Forest was trained on the training set, and the resampling statistics generated from the training process are given in Figure 7.4. The final

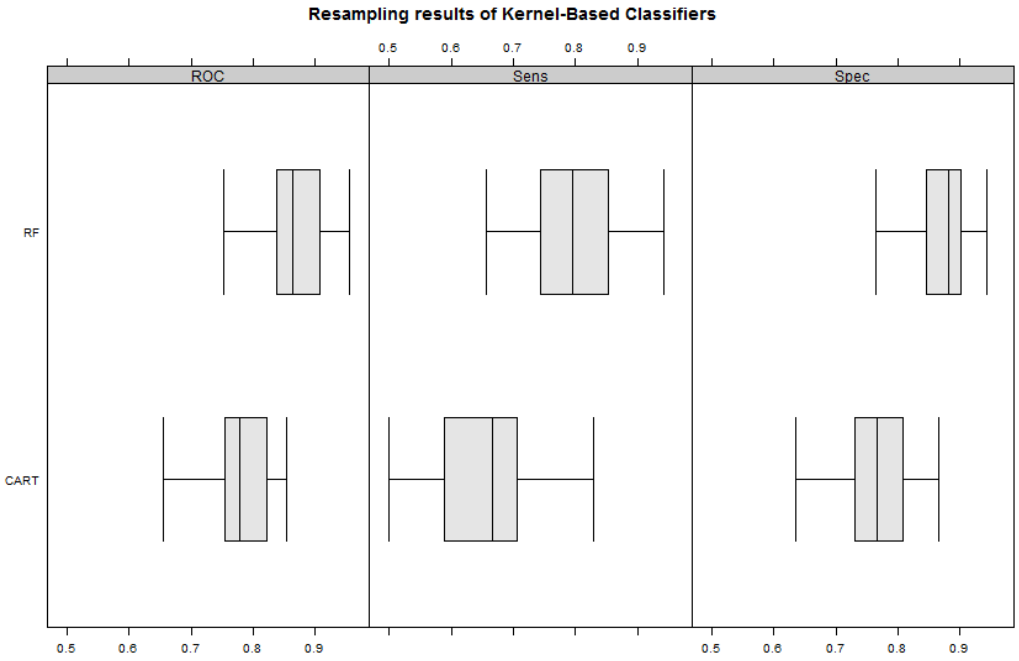

Figure 7.4: Resampling results of decision tree models showing the mean and deviance of the ROC area under curve, sensitivity and specificity

optimal fit for the Random Forest is with trees = 1 000, and *interaction.depth* = 3. The results obtained from training the models across the possible parameters is shown in Table A.4 on page 129. The ROC curve generated by fitting the final model to the testing data is included in Figure 7.7 at the end of this section. The area under the curve is equal to 0.893.

7.3.3 Function-based models

The classification models that rely on an activation function are evaluated in this subsection. The two models, in order of complexity, are logistic regression and neural networks. The results of the two models are compared with each other, and then at the end of this section compared with the models in the other categories.

Logistic functions are used in logistic regression to model how the probability of a prediction may be affected by the feature variables in the input data. The weighting functions need to be fitted to each input feature value in order to provide a linear function that describes the separation between the 'safe' and 'unsafe' classes. For this research, the implementation of logistic regression developed by Park and Hastie (2010) was used. The tuning variable for this model is lambda λ, the regularisation parameter for the L2 norm of the coefficients. This parameter λ introduces an additional cost to model complexity, in order to prevent the model from overtraining. The value

range tested for the tuning parameter during the construction of the model was $\lambda \in [10^{-4}, 10^{-1}]$. The results of training the model over this range of λ are shown in Table A.5 on page 130.

The logistic regression model considered for this research had an optimised value of $\lambda = 0.03$. This model's area under the ROC curve, sensitivity and specificity statistics are provided in the first row of Figure 7.5. The final model was applied to the hold-out testing set, and the ROC curve of the final performance is included in Figure 7.7 with a final value of the area under the ROC curve of 0.7313.

Neural networks represent an advanced form of function-based predictive modelling. A discussion about the theory of neural networks is provided in the Literature review, Chapter 2 of this dissertation, in Subsection 2.4.3. For this evaluation, the feed-forward neural network architecture was chosen as a representation of the neural network's possible performance on the ESD system. The implementation of neural networks in R chosen for this evaluation was previously implemented by Venables and Ripley (2002). The main tuning variable for the neural network under consideration is the number of neurons in the hidden layer. The range of hidden layer neurons considered is [5–20]. The results of training the neural network model over this range of nodes is shown in Table A.6 on page 130.

The optimised fit for the neural network is found to have 15 neurons in the hidden layer. Figure 7.5 shows the resampling statistics obtained from the

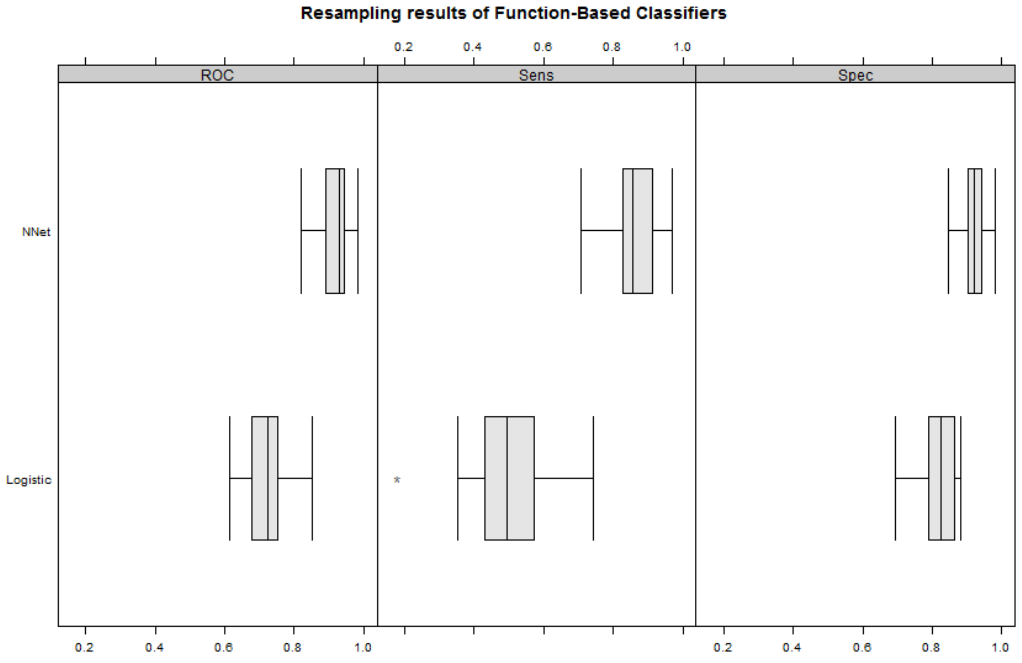

Figure 7.5: Resampling results of function-based models showing the mean and deviance of the ROC area under curve, sensitivity and specificity

k-fold cross-validation training process. Note the high sensitivity distribution observed during training. The model was then fitted to the testing set, and the ROC curve included in Figure 7.7 was obtained, with an area under curve value of 0.9329.

The neural network model implementation performed well, compared with the other classification models evaluated thus far. A high sensitivity value is seen in the second row of Figure 7.5, obtained from the k-fold cross-validation training process. To determine whether the same high level of performance is seen when the neural network is applied to the hold-out testing set, the ROC curve of the resulting values is included in Figure 7.7. In this figure, the total area under curve value is 0.933. This high value makes the neural network approach a strong contender for model optimisation, discussed in Section 7.4.

7.3.4 Kernel-based models

Support Vector Machines (SVMs) are evaluated in this section. SVMs are a powerful category of classification models that are extensively used in acoustic-related classification, as discussed in Subsection 2.4.4. This section implements and evaluates the two most common SVM kernels, the radial basis function (RBF) kernel and the polynomial kernel.

The first kernel evaluated is the radial basis function kernel, also known as the Gaussian kernel. The equation for the kernel is given by:

$$K(x, x') = \exp\left(-\frac{||x - x'||_2^2}{2\sigma^2}\right) \tag{7.3.1}$$

This kernel implies a tuning variable for the SVM model, the Gaussian-related sigma σ value. An additional tuning parameter is the regularisation C parameter which controls how much the separation boundary between classes can adapt to the smaller values in the training set.

The range of tuning variables chosen for the competitive model for the SVM with radial basis function is $\sigma = [0.01 : 1.0]$ and $C = [0.25 : 64.00]$. The implementation of the SVM model was done by Karatzoglou *et al.* (2004). Fitting the model on the training data with each of the tuning variables results in an optimised model with $\sigma = 0.0164$ and $C = 32$. The results of the training the model across the possible tuning parameters is given in Table A.7 on page 131. The area under the ROC curve, sensitivity and specificity statistics during resampling are shown in Figure 7.6. The fit of the optimised model to the testing set results in the ROC curve included in Figure 7.7, with an area under the ROC curve of 0.785.

The same general SVM implementation by Karatzoglou *et al.* (2004) is used again for the evaluation of the polynomial kernel on the ESD data. The polynomial kernel's equation is as follows:

$$K(x, y) = (x^T y + c)^d \tag{7.3.2}$$

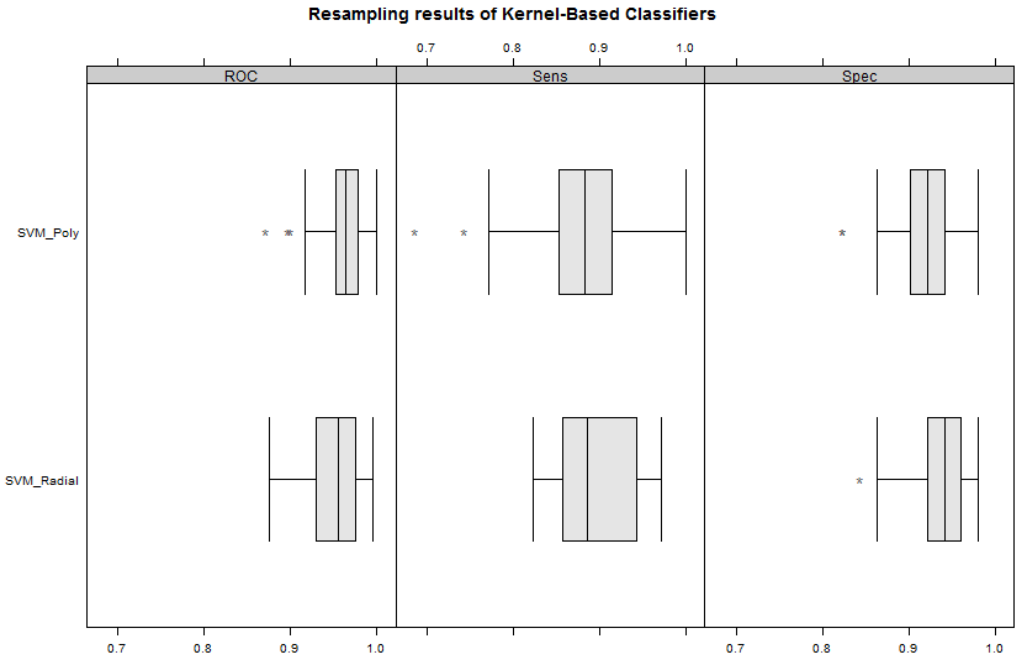

Figure 7.6: Resampling results of kernel-based models showing the mean and deviance of the ROC area under curve, sensitivity and specificity

From this equation, the tuning variables are as follows; $C \in [0.25, 0.50, 1.00]$ and $d \in [1, 2, 3]$. The results of training the model across the range of these variables are shown in Table A.8 on page 132. The optimal fit to the training data of the SVM with the polynomial kernel model is determined to be with the tuning parameters $C = 0.25$ and $d = 3$. The result of the resampling statistics can be seen in Figure 7.6. The trained model was applied to the testing set to estimate general performance, and a ROC curve was generated with an area under curve of 0.7983, which is included in Figure 7.7.

The results show that the two kernels have virtually the same response for the SVM model. The polynomial kernel slightly outperforms the RBF kernel in the area under ROC curve metric, though the RBF kernel has a higher tendency with the sensitivity statistic. The polynomial kernel is often used in acoustic applications such as Natural Language Processing (Chang *et al.*, 2010), which could be correlated to the slightly higher performance of this kernel on the ESD acoustic data.

7.3.5 Comparative evaluation of models

Important results that were omitted in the previous subsections are those obtained from training the models on the experimental dataset which was generated in Section 5.3 on page 57, and analysed in Chapter 6. The same model training method was used to determine the optimal models for use on the

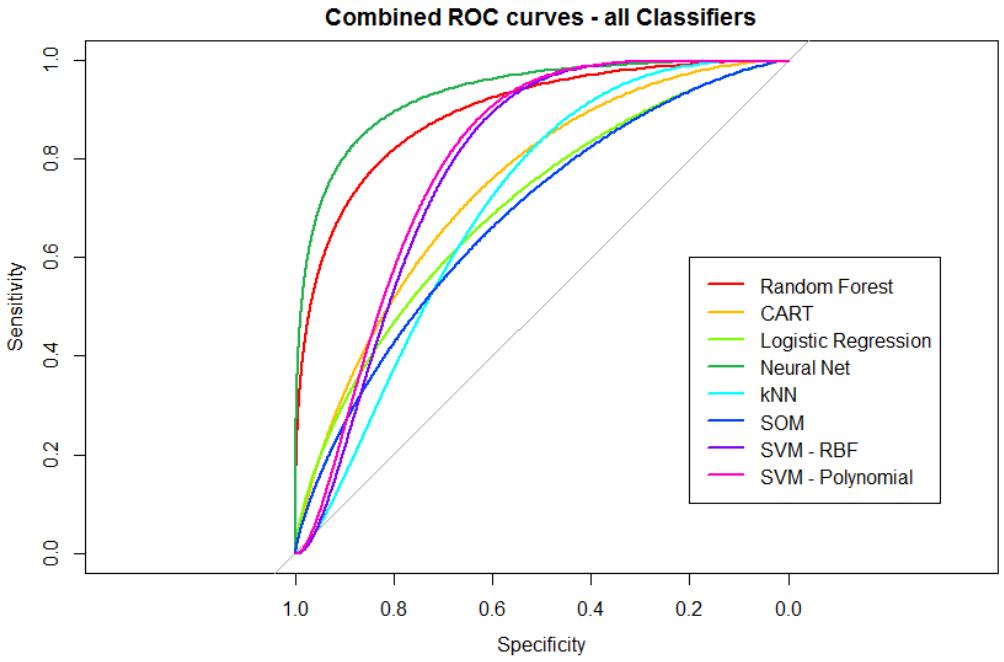

Figure 7.7: Combined ROC Curves of all the models competitively evaluated against each other

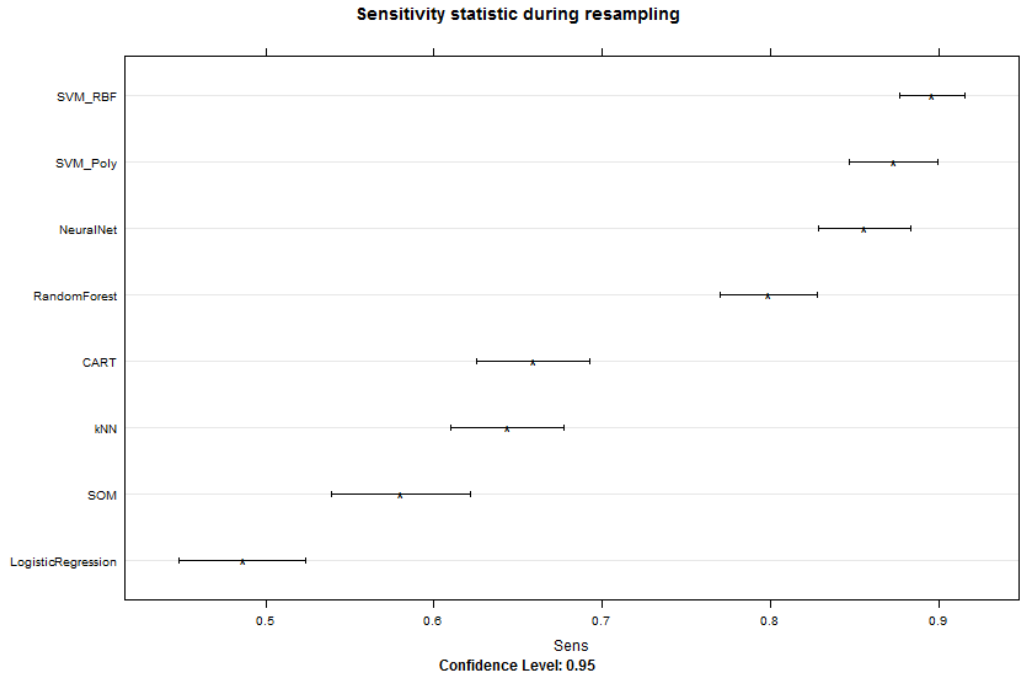

Figure 7.8: Sensitivity measure during resampling for all competitive models

experimental dataset. Perfect prediction by most of the classification models was shown when evaluating the respective hold-out testing sets. This result can be explained by an examination of boxplots in Figure 6.3 on page 71. Note that the spectral mean distribution is very clearly separated by class. A linear separation boundary can be visually determined at approximately 2 600 Hz of the S_c feature that would provide a clear distinction between the classes of the experimental dataset. The CART, Random Forest, logistic regression, neural network and both SVM models were able to determine this separation in the data, and therefore were able to predict with perfect accuracy. The kNN and SOM models each misclassified one prediction. Due to this clear separation in classes, the models obtained from training on the experimental dataset were not usable on the operational dataset which is non-linear in its class separation. This observation extends to the potential use of the models trained on laboratory data and deployed in operational conditions, where they are expected to perform very poorly.

The combined ROC curves of all the models trained on the operational dataset and evaluated in the previous subsections are shown in Figure 7.7 on the previous page. From this figure, it can be seen that the two models that seem to perform best on the hold-out ESD testing set are the feed-forward neural network implementation with 15 neurons in the hidden layer, followed by the Random Forest classifier. However, looking at the results from the sensitivity measure during training in Figure 7.8, it can be seen that the SVM models seem to have performed best in this important performance metric for the ESD problem. Recall from Equation 7.2.1 on page 98 that the sensitivity of a model is a measure of how often the model would misclassify a loose rock as safe, which is a result with potentially harmful consequence. From the evaluation of the confusion matrix of the SVM results, the sensitivity of the prediction on the testing set is determined to be 0.9062. This is higher than the specificity of 0.8140. This same relationship can be seen in Figure 7.7 where the SVM models are slanted towards the positive sensitivity side of the distribution, i.e both lines peak towards the upper edge of the plot.

The neural network model performs best in both the ROC area under curve measure, as well as one of the top performers in the sensitivity measure. This is an encouraging result, as neural networks perform well with noisy data, which is prevalent in the dataset as well as expected in operational conditions. The neural network has has a ROC AUC value of 0.933, a sensitivity metric of 0.883 and a specificity of 0.925. Note that the sensitivity lags behind that of the SVM models, though not necessarily statistically significantly so.

For this reason, the choice of models to evaluate further in the next section for optimal fit is the neural network, and the SVM model with the radial basis function (RBF) kernel.

7.4 Model optimisation through feature reduction

In this section, the best-performing classification models identified in the previous section are further refined. The competitive evaluation of models in the previous section already included a degree of optimisation in the form of tuning model parameters, but the idea of the choice of optimal model variables can be extended to other aspects of the training process.

The so-called 'curse of dimensionality' is a major concern in the field of classification models. The expression was created by Bellman (1956) to refer to the fact that many models work well in low dimensions but become intractable in higher dimensions. Generalising a model becomes exponentially harder as the number of features in a sample grows, because a fixed-size training set will cover a much smaller fraction of the possible input space.

Some classification models can also be negatively affected by the presence of features that are non-informative. Adding features of non-informative data to models and evaluating the resultant accuracy of prediction would add complexity to most models. An exception to this rule is the general category of tree-based classification models. CART and Random Forest already include a measure of dynamic feature elimination in the construction of the models. It is for this reason that feature reduction is presented at this point of the chapter, the only other strong contender for best model for this project was the Random Forest classifier and this discussion is not relevant to that model.

To judge which features are redundant, or even detrimental, to the effectiveness of the model a similar search algorithm can be deployed as that used in the previous section to determine the optimal model tuning parameters. Such methods conduct a search of the features to determine which produce the best results when entered into the model. There are three approaches exist for feature selection, as follows:

Forward selection: Features are added to the model one at a time, and the resultant prediction is evaluated for statistical significance.

Stepwise selection: Features are removed from the model individually, and the rest kept.

Backward selection: Features are iteratively removed from the model.

A variation on backward selection is called recursive feature elimination (RFE), which avoids training a model each time a feature is removed (Guyon and Elisseeff, 2003). A prerequisite to enable this elimination is determining which features contribute most to the prediction capability of a model.

Gevrey *et al.* (2003) use a combination of the absolute value of weights in neural network models to determine feature significance. The caret R package implements this method as a helper function. Assessing the significance of the features in the neural network model of this research results in the importance

of the individual features that are shown in Figure 7.9. By examining the results, it can be seen that *mel_bin9* and *mel_bin11* are considered the most important indicators to determine structural stability. These features correspond respectively to the energy in frequency bands [1 200 Hz–2 022 Hz] and [3 000 Hz–3 933 Hz]. These are similar frequency ranges to those that Hanson (1985) found to be effective for his loose rock acoustic assessment system, as discussed in Subsection 2.2.2 on page 13 of this dissertation.

It was found that reducing the existing features with the RFE method on the individual models did not result in a statistically significant improvement in the predictive results on the independent set for either of the classification models. The neural network had an initial ROC value of 0.933 for the full feature set, and 0.936 for a set with two features removed. The SVM model with the polynomial function had a full-set ROC value of 0.785, which increased to 0.793 with three features removed. Neither of these improvements is statistically significant.

An alternative considered for feature elimination in this project is a modification of the stepwise feature selection technique, based on the elimination of the specific feature groupings of this project. The three feature groupings derived in Chapter 6 and used in the combined feature set were the spectral descriptors, the frequency band content, and the MFCCs. The assumption was that each of these groupings of features contain information about the acoustic data that is useful and unique. However, by applying the modified stepwise feature selection technique to the data this assumption can be tested.

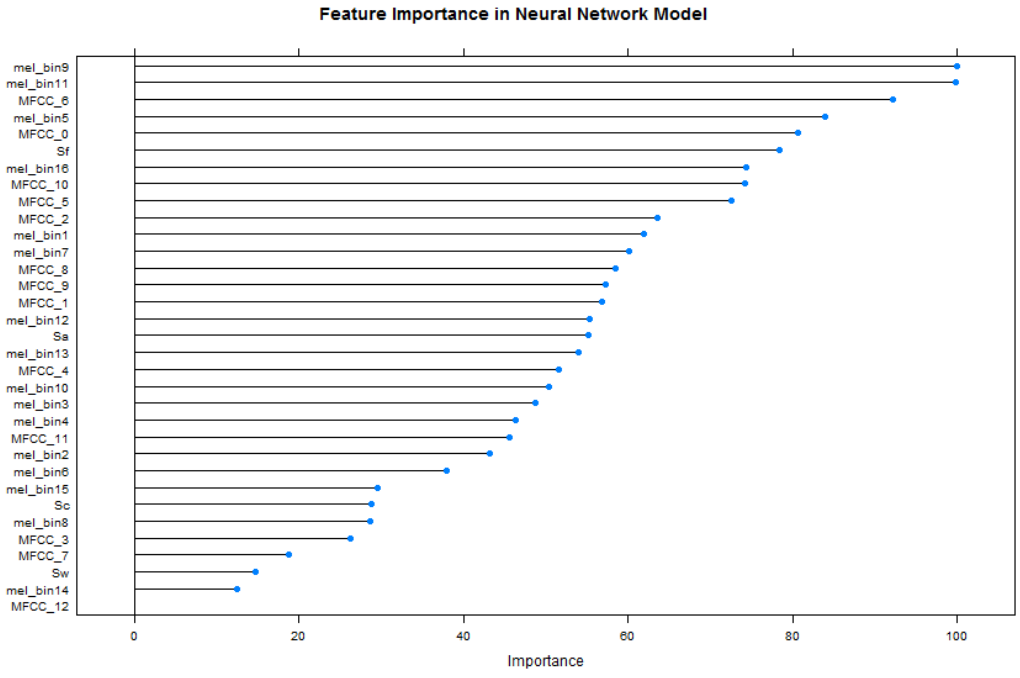

Figure 7.9: Feature significance for neural network model

Table 7.3 shows the results from applying the neural network model to reduced feature sets. Each of the possible convolutions of the groupings were used. The final result shows that the combination of all three feature groupings show the best performance, therefore the assumption that each grouping contributes to the structure of the combined feature group holds true.

Another way to reduce the feature dimensionality is to construct new features, which attempts to distil and retain the information in the existing features. This approach is called 'feature extraction'. Feature extraction aims to transform data in high-dimensional space to a space with fewer dimensions while retaining the variance of the data. Principal component analysis (PCA) is a multivariate technique that summarises systematic patterns of variation in the data by performing a linear mapping of the data. This technique attempts to transform observed features into a set of new features called the 'principal components', which are uncorrelated and reveal the dominant types of variations in the samples.

The steps in deriving the principal components are as follows:

1. Compute the mean for each feature (Subsection 6.6).
2. Compute the covariance values of all the features (Subsection 6.6.2).
3. Compute eigenvectors and then the corresponding eigenvalues.
4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a $d \times k$ dimensional matrix W.
5. Use this matrix to transform samples onto new subspace $y = W^T \times x$, where x is one sample.

The choice of how many principal components k to use is another search problem, though the variance of the individual components gives an indication of which ones represent the most information. Visualising the variance per principal component is known as a screeplot, and is shown in Figure 7.10.

The function *prcomp* is included in the base R installation, which means it is considered authoritative enough to form the basis of PCA calculations in R. This function was used to create candidate principal components from

Table 7.3: Feature grouping stepwise elimination for neural network model

Feature grouping tested	ROC	Sens	Spec
Descriptor	0.735	0.607	0.850
Band	0.790	0.673	0.856
MFCC	0.780	0.672	0.852
Descriptor + Band	0.831	0.707	0.858
Descriptor + MFCC	0.867	0.786	0.876
Band + MFCC	0.892	0.870	0.896
All combined	0.933	0.883	0.925

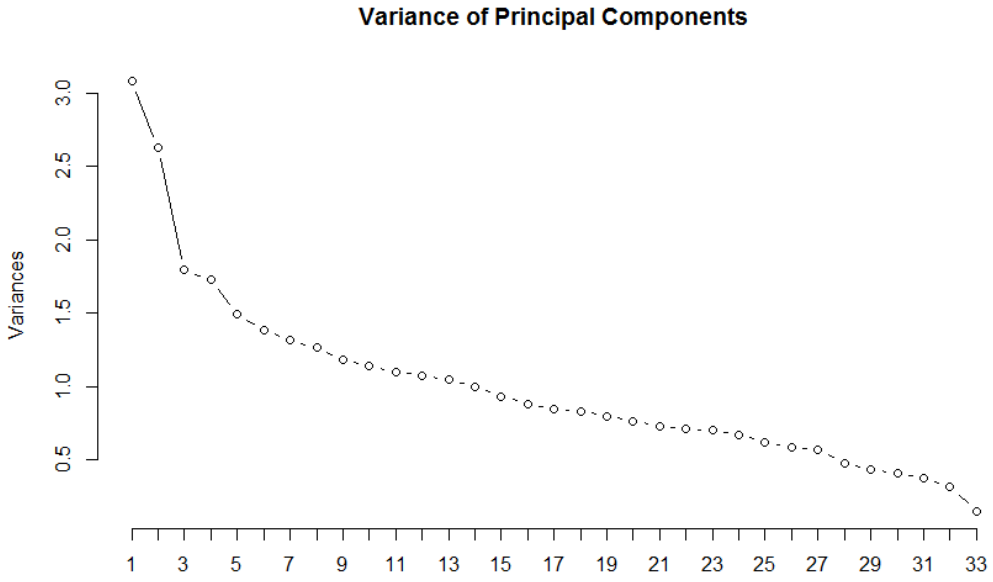

Figure 7.10: Descending variance of principal components

the ESD feature data. No non-informative components were found in the final transformation that could easily be discarded, so a threshold had to be implemented to discard components below certain thresholds.

Based on visual inspection of Figure 7.10, a drastically reduced subset of the data was considered. This subset consisted of the first four principal components. The performance of the SVM model showed significant degradation, from an ROC AUC value of 0.785 to 0.744. The neural network model showed a similar degradation, from 0.933 to 0.895, though the order of decrease is proportionally less significant than that of the SVM.

An alternative PCA reduction approach is suggested by Venables and Ripley (2002), who states that a standard choice is to discard principal components with a variance less than the standard deviation of the first component. After applying this criterion, a total of 29 components was left over. This is a reduction in dimensionality of 4 from the original 33 features.

The two contender classification models were trained and tested on the result. The SVM showed a significant increase in ROC AUC value from 0.785 to 0.813. However, the neural network still displayed the best performance with a ROC AUC value increase from 0.933 to 0.942. The final plot of the neural network with the reduced feature set is shown in Figure 7.11 on the facing page. The neural network model, like the other models evaluated in this project, outputs a single value on a scale between 0 and 1, indicating 'safe' and 'unsafe' respectively. To determine the interpretation of the value, a threshold boundary is applied. By examining the ROC curve, the optimal threshold

value that maximises both sensitivity and specificity can be determined. The optimal threshold value for the neural network model is found to be 0.483, which is indicated in Figure 7.11.

Based on this result, the final choice of classification model for the ESD is the neural network model with 15 hidden-layer neurons, and with a pre-processing PCA transformation of the feature set from 33 variables to 29 variables.

7.5 Model implementation on ESD

The final choice of model that is best suited for the ESD was identified in the previous section as the neural network model. This evaluation was done in the statistical programming language R. R is not suitable as an implementation language, and was not designed for that purpose. Neural network implementations are found in most programming languages, and the goal was to find one that functions well on the ESD.

The goal was to use the model specification and parameters generated in R. These specifically include the number of neurons in the hidden layer, the activation functions inside these neurons, and the values of the weights between the nodes. Furthermore, the W transformation matrix generated by the PCA needed to be applied as an intermediate function between the ESD feature

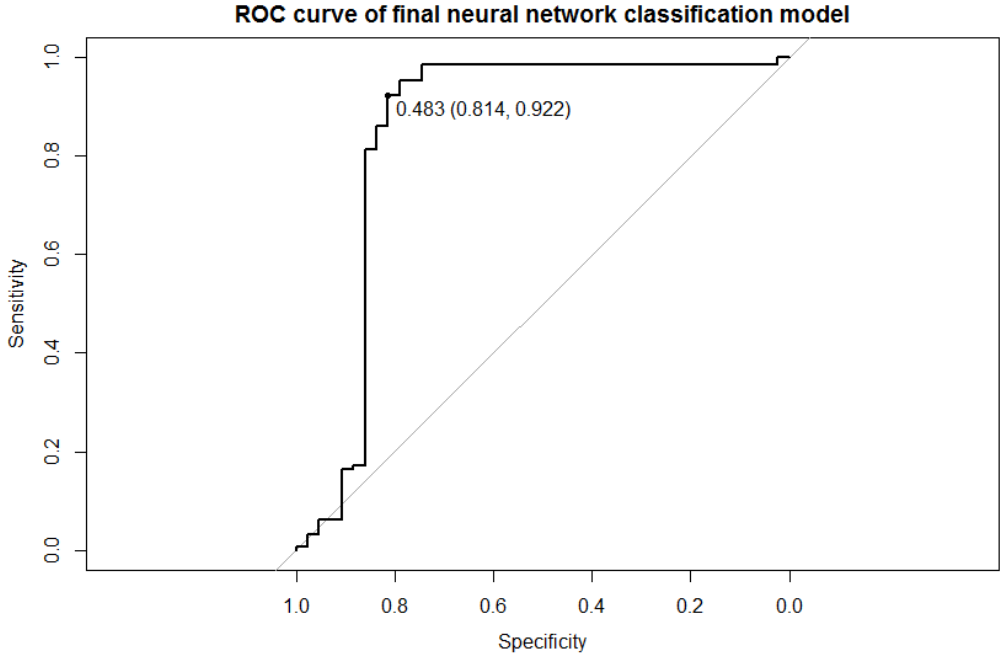

Figure 7.11: Final ROC plot of neural network model showing ideal threshold value to separate the output classes

input and the neural network trained on the PCA data.

The driver program on the ESD itself is written in `C`. This choice was made for interoperability with the codec chip-set that captures and encodes the acoustic waveforms, as well as for other hardware-related functions such as the driving of the LEDs and the power management on the device.

To implement this research on the ESD, the driver program was modified to include the functions necessary to transform the acoustic waveforms into the feature values as derived in Chapter 6. A function capable of discrete Fourier transform and discrete cosine transform was incorporated through the FFTW library in `C`, designed by Frigo and Johnson (2005). Functions capable of deriving the statistical moments of the distributions and the implementation of mel frequency range aggregation were self-coded in `C`.

The operating system on the Gumstix in the ESD is a Linux-based system. This enables the branching of new applications to be called from within a running program. Due to this freedom, the `C` driver program can activate any classification model application written in any other language, as long as it is able to provide the derived features to the program, as well as access the prediction result. This was implemented through the use of shared text files. The first file which was written by the driver program with the input features, after which the driver program activates the classification application. The program then monitors the classification application until it returns with a prediction result and terminates itself. The driver program reads the prediction result into a second text file.

NeuroSolutions is a software product by NeuroDimension that specializes in the construction, training and implementation of neural networks (Principe *et al.*, 2005). The practical use of this software for this dissertation is the ability of this product to generate modifiable `C++` code. Generating a neural network with the same specifications as were derived in this dissertation resulted in `C++` source code with an external text file with the parameters of the model. These parameters include the W matrix for the PCA transform, as well as the individual weights of the internode connections. Modifying this text file with this research's conclusion resulted in `C++` code that directly replicates the neural network model that was generated in R. Compiling the resultant source code for the ESD platform resulted in a single application that can be called from the ESD driver program.

NeuroSolutions provides a testbench module that can assess the constructed neural network model with a provided testing dataset. The testing set derived in Section 7.2.1 on page 96 consists of 134 samples that the classification model was not exposed to during training. The performance of the neural network model ported into NeuroSolutions can be derived by applying the NeuroSolution model on this test set and analysing the prediction results. The prediction results are given in Table 7.4 on the facing page. The performance of the model implemented on the ESD is derived from applying Equation 7.2.1 and Equation 7.2.2 on page 98 to the values presented in Table 7.2.1. The sensitivity of

this model is 92.94%, and the specificity is 85.71%. The specificity of the ESD model compared to the R-derived model is lower due to the presence of one additional false negative prediction. Variations in the model implementation algorithm between the R package and NeuroSolutions' implementation may account for this difference.

Table 7.4: Results from the NeuroSolutions implementation of the neural network model on the testing dataset

	Real 'unsafe'	Real 'safe'
Predicted 'unsafe'	79	7
Predicted 'safe'	6	42

7.6 Conclusion

This chapter shows the work done in creating a successful classification model for the ESD. The consistent testing environment was specified, which included the separation of the training data and the independent testing data, as well as the performance measure used in assessing model efficacy. Eight models were evaluated in this manner, from four distinct classification approaches. The best-performing models were further tested for optimisation by reducing the input feature dimensional space through the use of the PCA transformation.

The chosen model based on performance, a feed-forward neural network with a single hidden layer containing 15 neurons and a PCA pre-processing step, was specified with regard to its parameters. This model was then successfully implemented on the ESD in the `C++` programming language.

Chapter 8

Project Conclusion

This chapter provides a summary of the conclusions derived from the work in this dissertation. Its main purpose is to relate the outcome of the research in terms of the research questions and objective as stated in Chapter 1.

8.1 Summary of findings

The findings obtained from each chapter are presented in overview in this section. The results from the individual aspects of the dissertation inform the conclusions in the next section.

An investigation of the literature regarding the acoustic classification of structurally loose rocks during entry examination was presented in Chapter 2. It was found that devices capable of acoustic classification do exist, but that specific limitations in these devices prohibited their easy use during entry examination in deep South African hard rock mines. The classification devices evaluated in the literature either have a preparation phase that requires the device to be attached directly to the rock, or the devices are limited to strictly defined frequency ranges that were determined for rock assessment in shallow coal mines. Successful applications of classification models and feature transformations applicable to acoustic identification of structural weakness were found in the domain literature, and viable models for testing in this dissertation were identified.

A methodology was detailed in Chapter 3 that provided a process that could derive a justifiable result for this dissertation. This methodology provided discrete steps in a framework for the analysis and transformation of the acoustic data, as well as providing a rigorous evaluation of the classification models.

Chapter 4 presented the specification for the ESD. A modified version of the ESD that was used in this project for the collection of labelled acoustic samples was described.

The primary data used in this dissertation were the labelled acoustic samples recorded on the ESD during the entry-examination shifts in different working environments at Gold Fields' Driefontein mining operation. The process and data collected are detailed in Chapter 5. Teams consisting of mine engineers, shift supervisors and miners experienced in the process of entry examination and sounding for loose rocks recorded and labelled 710 acoustic samples of the sounding process for use in this research. In addition, 200 samples were collected from a controlled testing environment to provide an idealised set of data as a control group.

The acoustic data collected were analysed and transformed in Chapter 6. The acoustic samples were transformed into three feature groups applicable to classification models. These three groups were respectively spectral descriptors, frequency band content and MFCCs. The feature data were analysed for insight into the acoustic data collected, and applicable data cleaning of the feature set was performed. This resulted in a set of 513 samples for use in training and testing classification models. Comparative evaluation between individual features was performed to test for correlation of features in the data set. Lastly, the individual features were scaled to the same range and variance to enable their effective use as inputs to classification models.

The identification of the most suitable classification model was done in Chapter 7. Eight classification models from four separate classification approaches were competitively evaluated against each other in order to find the model best suited for this research. The input feature data were divided into a training set and an independent testing set. Classification models were trained on the same training set in a k-fold cross-validation algorithm, and optimised tuning parameters were found for each model. Models were compared on their sensitivity performance, as measured by area under ROC curves. A neural network classification model was identified as the optimal model for the ESD due to its high level of accuracy and sensitivity. Final optimisation of the model was done by reducing the input feature dimensionality by means of pre-processing the features through the method of PCA.

8.2 Conclusions

In the introduction to this dissertation, the research goals and final research objective for this dissertation were stated. This section will review these goals in terms of the results shown in this dissertation.

Gathered a dataset of labelled acoustic data from the sounding process in a mine-wide geotechnical area: The CSIR and Gold Fields partnered to collected acoustic waveform samples from nine working stopes representing four distinct reefs at the Driefontein mining operation. Samples were collected by experts in the entry-examination procedure in teams, represent-

ing various levels of mining personnel and knowledge levels. Each sample was labelled by means of a button input at the time of collection, so that each acoustic waveform file on the ESD had a matched label. Documentation was kept on the environmental conditions at the collection sites, including observations about the ground, water and ambient noise conditions.

Derived a descriptive representation of the acoustic data for the use in classification models: A descriptive representation of each acoustic sample was obtained by transforming the raw acoustic waveforms into discrete features. Three feature approaches were identified and implemented: spectral distribution descriptors, frequency band energy content, and MFCCs. These features were corroborated with sources in the domain literature. Evaluation in this dissertation also showed that each of the 33 individual features had information content that assisted the classification models to make more accurate predictions. A slight reduction in the dimensionality of the feature set was found to increase effectiveness in the neural network model. This was shown by an additional PCA transformation of the features to expose maximum variance in the individual features, and a total of 29 final inputs was used.

An observation from the evaluation of the features in relation to classification model outputs showed that the frequency bands that are most significant at differentiating a loose rock from a solidly attached rock are [1 200 Hz–2 022 Hz] and [3 000 Hz–3 933 Hz].

Found an effective classification model for the judgement of rock integrity: A neural network classification model was found to be most effective at predicting loose rocks with the ESD. The parameters of the neural network are as follows:

- Single hidden layer with 15 neurons
- Sigmoid activation function
- PCA preprocessing, from 33 features to 29 principal components
- One output between 0 and 1 indicating the probability of an 'unsafe' rock state, with a rounding threshold at 0.483.
- Values for the inter-neuron weights representing the trained version of the model, as well as the transformation matrix W for the transformation of feature inputs into the principal components.

This model's performance on an independent testing set that it had not been exposed to during the training of the model, shows the following performance statistics. Note that the balanced accuracy of the model consists of the arithmetic mean of the model's sensitivity and specificity.

Model sensitivity: 92,94 %

Model specificity: 87,76 %
Model balanced accuracy: 90,35 %

Implemented the classification model successfully on the ESD: The neural network model was integrated as a separate application to the main ESD programming. The ESD programming consisted primarily of a driver program written in `C`. This driver program was written to capture the acoustic waveforms and transform them into the feature values used by the classification model. The model application is implemented as a `C++` program, based on an implementation generated in NeuroSolutions modified by the optimal fit parameters generated in the R statistical programming language. The driver program calls up the model each time an acoustic event is captured, provides the model with the feature values of the event, and reads back the prediction results from the model.

Achieved the research objective: The overall research objective for the project described in this dissertation is as follows:

The development of an effective classification system that is able to determine the stability of rock hanging walls in underground mining environments during entry examination based on the acoustic signal generated in the sounding process

The conclusions listed for the achievement of each of the previous research goals successfully realise this objective.

8.3 Summary of contributions

The combination of the acoustic data recorded for this research, the features derived from the data, and the classification model developed is a novel addition to the problem of comprehensive entry-examination. Due to the fact that the chosen classification model effective performs purely on acoustic input, it provides assistance in identifying loose rocks without being an intrusive part of the miner's procedure. This differs from other approaches made in previous research, where physical connection of the instrument to the rock was a prerequisite and the frequency ranges evaluated were limited and fixed.

It is hoped that the implementation of an effective classification system on the ESD will spur the deployment of the device in the mining industry. If the device can contribute to the increased detection of hazardous loose rocks, it will make a direct contribution to the management of risk in underground working environments.

8.4 Suggestions for further research

The next version of the ESD can be designed with an interactive visual user interface. The new version can be designed with improvements identified in this research. The new design can incorporate the judgement from multiple strikes on the same rock, and base its classification on a majority vote of the derived predictions. Real-time adjustments to the model can also be incorporated through the improved interface. Samples encountered during normal operation can be stored for further training and optimisation of the classification system. Anomalous samples can be discarded when they are encountered.

Further research related to the topic of this dissertation could stem from variations of the methodology used in this project. Particularly, the choice of feature representation used in this research, as well as the choice and method of raw data collected, could be expanded. This concluding section briefly suggests expansion on these ideas, and possible further research that could stem from them.

The representation of acoustic analysis in this research was purely in the frequency domain. There may be additional acoustic features can exist in the time-domain representation that could enable better performance of the classification model. A specific feature transformation that represents both time and frequency characteristics is that of the discrete wavelet transform (DWT). Wavelet length determines the resolution of the acoustic feature being transformed: short lengths give an indication of short temporal changes, whereas a stretched wavelet provides signal-wide spectral characteristics. Additional research in this regard could justify the inclusion of wavelets in acoustic structural assessments in underground environments.

This dissertation focused exclusively on the use of acoustic waveform inputs for the determination of hazardous rock conditions. However, other potentially detectable physical phenomena could be used in conjunction with acoustic data for this purpose. Thermal profiling of rocks in deep underground environments has been shown to be useful in visually delineating rocks that are loosely coupled to the surrounding rock mass. The use of visual pattern recognition from a thermal camera in a combined classification system with the ESD could deliver an improved analysis of the safety of the rock.

The acoustic data obtained in this research were generated exclusively through the use of pinch bars wielded by miners. The use of a pinch bar for striking rock results in variable impact energies and strike angles. Research into an actuating instrument to replace the pinch bar could be fruitful. A classification model could perform better if the rock was struck in a more consistent fashion through this method. A practical extension of this actuator would be its usability on a mobile autonomous platform such as a robot system deployed in a mining environment. This would allow aspects of entry examination to be carried out without a person being exposed to the risk of rock falls.

Appendices

Appendix A

Model Tuning Results

The intermediate results obtained from varying the classification model tuning parameters are given in this appendix. For each model, the optimised combination of parameters is determined by the resulting area under ROC curve value, in combination with its peak sensitivity and specificity performance on the operational dataset at the optimal threshold. The choice of parameters made for each model is presented in Chapter 7.

Table A.1: kNN classification model tuning parameters values tested and resultant statistics

k	ROC	Sens	Spec
2	0.744	0.652	0.764
5	0.752	0.643	0.788
7	0.731	0.559	0.723
9	0.723	0.582	0.759
11	0.708	0.542	0.741
13	0.689	0.530	0.757
15	0.691	0.524	0.754
17	0.678	0.530	0.756
19	0.688	0.519	0.774
21	0.685	0.515	0.774

Table A.2: SOM classification model tuning parameters values tested and resultant statistics

xdim	*ydim*	*xweight*	ROC	Sens	Spec
1	2	0.5	0.616	0.594	0.681
1	2	0.7	0.608	0.592	0.667
1	2	0.9	0.597	0.566	0.682
1	3	0.5	0.516	0.557	0.662
1	3	0.7	0.500	0.529	0.669
1	3	0.9	0.519	0.475	0.743
1	4	0.5	0.552	0.514	0.702
1	4	0.7	0.560	0.533	0.679
1	4	0.9	0.569	0.537	0.693
2	2	0.5	0.551	0.524	0.701
2	2	0.7	0.553	0.494	0.727
2	2	0.9	0.595	0.533	0.701
2	3	0.5	0.610	0.449	0.818
2	3	0.7	0.621	0.416	0.833
2	3	0.9	0.608	0.465	0.805
2	4	0.5	0.633	0.570	0.700
2	4	0.7	0.639	0.584	0.698
2	4	0.9	0.631	0.590	0.706
3	3	0.5	0.663	0.520	0.772
3	3	0.7	0.659	0.513	0.776
3	3	0.9	0.645	0.511	0.792
3	4	0.5	0.687	0.580	0.744
3	4	0.7	0.672	0.546	0.761
3	4	0.9	0.664	0.543	0.774

Table A.3: CART classification model tuning parameters values tested and resultant statistics

cp	ROC	Sens	Spec
0.00000	0.775	0.659	0.770
0.00777	0.751	0.628	0.785
0.01555	0.720	0.576	0.817
0.02332	0.720	0.587	0.786
0.03109	0.657	0.542	0.803
0.03887	0.485	0.460	0.831
0.04664	0.420	0.444	0.842
0.05441	0.420	0.444	0.842
0.06219	0.420	0.444	0.842
0.06996	0.420	0.444	0.842
0.07774	0.420	0.444	0.842
0.08551	0.420	0.444	0.842
0.09328	0.420	0.444	0.842
0.10106	0.420	0.444	0.842

Table A.4: Random Forest classification model tuning parameters values tested and resultant statistics

interaction.depth	n.trees	ROC	Sens	Spec
1	250	0.721	0.379	0.886
1	500	0.737	0.435	0.872
1	1000	0.762	0.489	0.864
1	1750	0.775	0.539	0.846
1	2000	0.781	0.552	0.846
3	250	0.789	0.494	0.874
3	500	0.820	0.572	0.871
3	1000	0.941	0.867	0.938
3	1750	0.938	0.866	0.938
3	2000	0.916	0.769	0.915
5	250	0.864	0.650	0.896
5	500	0.820	0.558	0.880
5	1000	0.913	0.785	0.911
5	1750	0.903	0.755	0.904
5	2000	0.866	0.668	0.879

Table A.5: Logistic regression classification model tuning parameters values tested and resultant statistics

λ	ROC	Sens	Spec
0.0000	0.622	0.416	0.833
0.0001	0.643	0.465	0.805
0.0003	0.667	0.570	0.700
0.001	0.702	0.584	0.698
0.003	0.717	0.513	0.706
0.01	0.735	0.520	0.772
0.03	0.741	0.590	0.776
0.10	0.712	0.511	0.792

Table A.6: Neural network classification model tuning parameters values tested and resultant statistics

size	ROC	Sens	Spec
5	0.839	0.757	0.868
8	0.886	0.839	0.901
10	0.903	0.847	0.909
13	0.923	0.873	0.920
15	0.934	0.883	0.925
18	0.926	0.870	0.920
20	0.903	0.844	0.898

Table A.7: SVM with RBF kernel classification model tuning parameters values tested and resultant statistics

C	sigma	ROC	Sens	Spec
0.25	0.0082	0.722	0.579	0.806
0.25	0.0164	0.751	0.565	0.837
0.25	0.0328	0.809	0.627	0.838
0.50	0.0082	0.739	0.518	0.850
0.50	0.0164	0.780	0.572	0.852
0.50	0.0328	0.852	0.637	0.859
1.00	0.0082	0.756	0.550	0.852
1.00	0.0164	0.820	0.601	0.855
1.00	0.0328	0.905	0.752	0.880
2.00	0.0082	0.785	0.570	0.849
2.00	0.0164	0.865	0.668	0.870
2.00	0.0328	0.949	0.844	0.913
4.00	0.0082	0.831	0.607	0.858
4.00	0.0164	0.909	0.770	0.896
4.00	0.0328	0.945	0.875	0.930
8.00	0.0082	0.867	0.686	0.876
8.00	0.0164	0.937	0.838	0.908
8.00	0.0328	0.965	0.876	0.935
16.00	0.0082	0.903	0.766	0.896
16.00	0.0164	0.943	0.872	0.932
16.00	0.0328	0.962	0.875	0.933
32.00	0.0082	0.924	0.818	0.894
32.00	0.0164	0.965	0.876	0.933
32.00	0.0328	0.951	0.871	0.932
64.00	0.0082	0.929	0.852	0.919
64.00	0.0164	0.945	0.879	0.927
64.00	0.0328	0.956	0.879	0.934

Table A.8: SVM with polynomial kernel classification model tuning parameters values tested and resultant statistics

degree	C	ROC	Sens	Spec
1	0.25	0.708	0.553	0.800
1	0.50	0.714	0.489	0.848
1	1.00	0.714	0.477	0.855
2	0.25	0.726	0.500	0.852
2	0.50	0.735	0.507	0.850
2	1.00	0.823	0.616	0.861
3	0.25	0.894	0.787	0.901
3	0.50	0.882	0.717	0.875
3	1.00	0.839	0.618	0.857
4	0.25	0.873	0.698	0.872
4	0.50	0.834	0.615	0.860
4	1.00	0.790	0.573	0.856
5	0.25	0.819	0.601	0.853
5	0.50	0.775	0.559	0.852
5	1.00	0.750	0.530	0.851

References

(1996). *Mine Health and Safety Act, 1996 (Act No. 29 of 1996).* Available at: `http://www.dmr.gov.za/legislation/`

Adams, G., Jager, A. *et al.* (1980). Petroscopic observations of rock fracturing ahead of stope faces in deep-level gold mines. *Journal of the South African Institute of Mining and Metallurgy*, vol. 80, no. 6, pp. 204–209.

Aggelis, D.G. (2011). Classification of cracking mode in concrete by acoustic emission parameters. *Mechanics Research Communications*, vol. 38, no. 3, pp. 153–157.

Allison, H. and Lama, R.D. (1979). Low frequency sounding technique for predicting progressive failure of rock. *Rock Mechanics and Rock Engineering*, vol. 12, pp. 79–97.

Altman, D.G. and Bland, J.M. (1994). Diagnostic tests 3: receiver operating characteristic plots. *BMJ: British Medical Journal*, vol. 309, no. 6948, p. 188.

Altman, N.S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, vol. 46, no. 3, pp. 175–185.

Bellman, R. (1956). Dynamic programming and Lagrange multipliers. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 42, no. 10, p. 767.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305.

Bosman, J., Malan, D. and Drescher, K. (2000). Time-dependent tunnel deformation at hartebeestfontein mine. *JOURNAL-SOUTH AFRICAN INSTITUTE OF MINING AND METALLURGY*, vol. 100, no. 6, pp. 333–340.

Breiman, L. (1999). Random forests–random features. *UC Berkeley, Statistics Department, Technical Report*, , no. 567.

Breiman, L., Friedman, J., Stone, C.J. and Olshen, R.A. (1984). *Classification and regression trees*.

Carceles, S.M. and Muller, D. (2007 June 19). Artificial intelligence systems for classifying events, objects and situations. US Patent 7,233,936.

Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M. and Lin, C.-J. (2010). Training and testing low-degree polynomial data mappings via linear SVM. *Journal of Machine Learning Research*, vol. 11, pp. 1471–1490.

Chiang, L.H., Braatz, R.D. and Russell, E.L. (2001). *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media.

Childers, D.G., Skinner, D.P. and Kemerait, R.C. (1977). The cepstrum: A guide to processing. *Proceedings of the IEEE*, vol. 65, no. 10, pp. 1428–1443.

Chu, S., Narayanan, S. and Kuo, C.-C. (2009). Environmental sound recognition with time–frequency audio features. *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 6, pp. 1142–1158.

Denkhaus, H., Hill, F. and Roux, A. (1958). A review of recent research into rockbursts and strata movement in deep-level mining in south africa. *Ass. Min. Mngrs. S. Afr*, pp. 245–268.

Dhanalakshmi, P., Palanivel, S. and Ramalingam, V. (2009). Classification of audio signals using SVM and RBFNN. *Expert Systems with Applications*, vol. 36, no. 3, pp. 6069–6075.

Dittmar, T. (2011). *Audio Engineering 101: A beginner's guide to music production*. ISBN 9780240819150.

Draxin (2014). Rock integrity monitoring solution.
Available at: `http://www.draxin.co.za/rock-integrity-monitor.html`

Fant, G. (1949). *Analys av de svenska konsonantljuden: talets allmänna svängningsstruktur*. LM Ericsson.

Fraga-Silva, T., Gauvain, J. and Lamel, L. (2011). Lattice-based unsupervised acoustic model training. In: *Preoceedings of the Acoustics, Speech and Signal Processing, ICASSP, 2011, IEEE International Conference*, pp. 4656–4659.

Frigo, M. and Johnson, S.G. (2005). The design and implementation of FFTW3. *Proceedings of the IEEE*, vol. 93, no. 2, pp. 216–231.

Gevrey, M., Dimopoulos, I. and Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, vol. 160, no. 3, pp. 249–264.

Gil, D. and Johnsson, M. (2010). Supervised som based architecture versus multilayer perceptron and rbf networks. In: *Proceedings of the Linköping Electronic Conference*, pp. 15–24.

Gillet, O. and Richard, G. (2004). Automatic transcription of drum loops. In: *Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing, ICASSP'04*, vol. 4, pp. iv–269. IEEE.

Gladwin, M.T. and Stacey, F. (1974). Anelastic degradation of acoustic pulses in rock. *Physics of the Earth and Planetary Interiors*, vol. 8, no. 4, pp. 332–336.

Green, J., Bosscha, P., Candy, L., Hlophe, K., Coetzee, S. and Brink, S. (2010). Can a robot improve mine safety?
Available at: `http://hdl.handle.net/10204/5022`

Grey, J.M. and Gordon, J.W. (1978). Perceptual effects of spectral modifications on musical timbres. *The Journal of the Acoustical Society of America*, vol. 63, no. 5, pp. 1493–1500.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, vol. 3, pp. 1157–1182.

Hannula, M., Laitinen, J. and Alasaarela, E. (2003). Classification accuracy of a frequency analysis method: Comparison between supervised SOM and KNN. In: *Information Technology Applications in Biomedicine, 4th International IEEE EMBS Special Topic Conference*, pp. 254–257.

Hanson, D. (1985). Rock stability analysis by acoustic spectroscopy. *Mining Engineering (Littleton, CO)*, vol. 37, pp. 54–60.

Hanson, D. (1986 07). Detached rock evaluation device. US Patent 4598588.
Available at: `http://www.patentlens.net/patentlens/patent/US_4598588/en/`

Hartman, H.L. and Mutmansky, J.M. (2002). *Introductory mining engineering*. New York, Wiley.

Janert, P.K. (2010). *Data analysis with open source tools*. O'Reilly.

Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A. (2004). kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20.
Available at: `http://www.jstatsoft.org/v11/i09/`

Khai-Wern, L., Tien, T.L. and Lateh, H. (2011). Landslide hazard mapping of penang island using probabilistic methods and logistic regression. In: *Proceedings of the IEEE International Conference on Imaging Systems and Techniques*, pp. 273–278.

Koenig, W. (1949). A new frequency scale for acoustic measurements. *Bell Telephone Laboratory Record*, vol. 27, pp. 299–301.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, vol. 43, no. 1, pp. 59–69.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, vol. 28, no. 5, pp. 1–26.

Kumar, B.R., Vardhan, H., Govindaraj, M. and Saraswathi, S.P. (2013). Artificial neural network model for prediction of rock properties from sound level produced during drilling. *Geomechanics and Geoengineering*, vol. 8, no. 1, pp. 53–61.

Lamel, L., Gauvain, J.-L. and Adda, G. (2002). Lightly supervised and unsupervised acoustic model training. *Computer Speech & Language*, vol. 16, no. 1, pp. 115–129.

Lee Rodgers, J. and Nicewander, W.A. (1988). Thirteen ways to look at the correlation coefficient. *The American Statistician*, vol. 42, no. 1, pp. 59–66.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomforest. *R News*, vol. 2, no. 3, pp. 18–22.
Available at: `http://CRAN.R-project.org/doc/Rnews/`

Lockner, D. (1993). The role of acoustic emission in the study of rock fracture. In: *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 30, pp. 883–899.

Ludeña-Choez, J. and Gallardo-Antolín, A. (2014). Feature extraction based on the high-pass filtering of audio signals for acoustic event classification. *Computer Speech & Language*, vol. 28, no. 1, pp. 93–107.

Lux, E.G., Mason, V.A., Pettigrew, M.J. and Rhodes, D.B. (1991 June 18). Loose rock detector. US Patent 5,024,090.

Ma, L., Milner, B. and Smith, D. (2006). Acoustic environment classification. *ACM Transactions on Speech and Language Processing (TSLP)*, vol. 3, no. 2, pp. 1–22.

Maillet, E., Godin, N., R'Mili, M., Reynaud, P., Fantozzi, G. and Lamon, J. (2014). Damage monitoring and identification in sic/sic minicomposites using combined acousto-ultrasonics and acoustic emission. *Composites Part A: Applied Science and Manufacturing*, vol. 57, pp. 8–15.

Makhoul, J. and Cosell, L. (1976). Lpcw: An lpc vocoder with linear predictive spectral warping. In: *Proceedings of the Acoustics, Speech, and Signal Processing, IEEE International Conference, ICASSP'76.*, vol. 1, pp. 466–469.

Malan, D. (1999). Time-dependent behaviour of deep level tabular excavations in hard rock. *Rock Mechanics and Rock Engineering*, vol. 32, no. 2, pp. 123–155.

McCrory, J.P., Al-Jumaili, S.K., Crivelli, D., Pearson, M.R., Eaton, M.J., Featherston, C.A., Guagliano, M., Holford, K.M. and Pullin, R. (2014). Damage classification in carbon fibre composites using acoustic emission: A comparison of three techniques. *Composites Part B: Engineering*, vol. 68, pp. 424–430.

Mierswa, I. and Morik, K. (2005). Automatic feature extraction for classifying audio data. *Machine learning*, vol. 58, no. 2-3, pp. 127–149.

Minsky, M.L. and Papert, S.A. (1987). *Perceptrons - Expanded Edition: An Introduction to Computational Geometry*. MIT press Boston.

Moerchen, F., Mierswa, I. and Ultsch, A. (2006). Understandable models of music collections based on exhaustive feature generation with temporal statistics. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 882–891.

Moerchen, F., Ultsch, A., Thies, M., Loehken, I., Nöcker, M., Stamm, C., Efthymiou, N. and Kümmerer, M. (2004). Musicminer: Visualizing perceptual distances of music as topograpical maps. Tech. Rep., Technical report, Dept. of Mathematics and Computer Science, University of Marburg, Germany.

Moore, B.C. and Glasberg, B.R. (1996). A revision of zwicker's loudness model. *Acta Acustica united with Acustica*, vol. 82, no. 2, pp. 335–345.

Naidoo, B. (2007 10). Tautona to take deepest mine accolade. Online.
Available at: `http://www.miningweekly.co.za/article_id=98516`

Novotney, S., Schwartz, R. and Ma, J. (2009). Unsupervised acoustic and language model training with small amounts of labelled data. In: *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 4297–4300.

Ntalampiras, S. and Roveri, M. (2013). Rock collapse forecasting: A novel approach based on the classification of micro-acoustic signals in the wavelet domain. In: *Sensors, 2013 IEEE*, pp. 1–4. IEEE.

Nyareli, T., Brink, V.Z. and Bilgeri, R. (2009). Development of an electronic sounding device. Internal Report CSIR/NRE/ER/2006/0199/A.

Park, M.Y. and Hastie, T. (2010). *stepPlr: L2 penalized logistic regression with a stepwise variable selection*. R package version 0.92.
Available at: `http://CRAN.R-project.org/package=stepPlr`

Pickles, J.O. (1988). *An introduction to the physiology of hearing*, vol. 2.

Platmine (2009). Presentation for the gauteng branch of sanire. Online.
Available at: `http://goo.gl/RBlui3`

Poggenpoel, J. (2012 September). SWP 10.06.03 - Entry Procedures Conventional Mining. SWP, Impala Rustenburg Operations.

Poria, S., Gelbukh, A., Hussain, A., Bandyopadhyay, S. and Howard, N. (2013). Music genre classification: A semi-supervised approach. In: *Pattern Recognition*, pp. 254–263.

Principe, J., Lefebvre, C., Lynn, G., Fancourt, C. and Wooten, D. (2005). Neurosolutions–documentation. *NeuroDimension, Inc.: Gainesville, FL*.

Quinlan, J.R. (1993). *C4.5: programs for machine learning*, vol. 1.

Ripley, B. and Venables, W. (2013). Package 'class'.
Available at: `http://r.meteo.uni.wroc.pl/web/packages/cllass/class.df`

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1985). Learning internal representations by error propagation. Tech. Rep., DTIC Document.

Schultz, R.L., De Jesus, O. and Osborne Jr, A.J. (2008 April 15). Method and apparatus for monitoring the condition of a downhole drill bit, and communicating the condition to the surface. US Patent 7,357,197.

Schweitzer, J. and Johnson, R. (1997). Geotechnical classification of deep and ultra-deep witwatersrand mining areas, south africa. *Mineralium Deposita*, vol. 32, no. 4, pp. 335–348.

Sorensen, H.V., Jones, D.L., Heideman, M. and Burrus, C.S. (1987). Real-valued fast fourier transform algorithms. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, no. 6, pp. 849–863.

Stevens, S.S. and Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *The American Journal of Psychology*, vol. 53, no. 3, pp. 329–353.

Stevens, S.S., Volkmann, J. and Newman, E.B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America*, vol. 8, p. 185.

Summerfield, P. (1956). *A study of the air and rock vibrations produced by impact testing of mine roof*, vol. 5251.

Takeshima, H., Suzuki, Y., Fujii, H., Kumagai, M., Ashihara, K., Fujimori, T. and Sone, T. (2001). Equal-loudness contours measured by the randomized maximum likelihood sequential procedure. *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 389–399.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, vol. 49, no. 8, pp. 1259–1266.

Therneau, T., Atkinson, B. and Ripley, B. (2014). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-8.
Available at: `http://CRAN.R-project.org/package=rpart`

Tibaduiza, D.-A., Torres-Arredondo, M.-A., Mujica, L., Rodellar, J. and Fritzen, C.-P. (2013). A study of two unsupervised data driven statistical methodologies for detecting and classifying damages in structural health monitoring. *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 467–484.

Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *Speech and Audio Processing, IEEE transactions on*, vol. 10, no. 5, pp. 293–302.

Umesh, S., Cohen, L. and Nelson, D. (1999). Fitting the mel scale. In: *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 1, pp. 217–220. IEEE.

Venables, W.N. and Ripley, B.D. (2002). *Modern Applied Statistics with S*. 4th edn. New York. ISBN 0-387-95457-0.
Available at: `http://www.stats.ox.ac.uk/pub/MASS4`

Wehrens, R. and Buydens, L. (2007). Self- and super-organising maps in r: the kohonen package. *J. Stat. Softw.*, vol. 21, no. 5.
Available at: `http://www.jstatsoft.org/v21/i05`

Wolpert, D.H. and Macready, W.G. (1997). No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on*, vol. 1, no. 1, pp. 67–82.

Wu, C., Jia, R. and Qiu, T. (2013). Rock burst monitoring and early warning based on incremental learning method with svm. *Research Journal of Information Technology*, vol. 5, no. 4, pp. 121–124.

Young, R.W. (1939). Terminology for logarithmic frequency units. *The Journal of the Acoustical Society of America*, vol. 11, p. 134.

Zwicker, E. and Fastl, H. (1999). *Psychoacoustics: Facts and models*, vol. 2.

Zwicker, E., Flottorp, G. and Stevens, S.S. (1957). Critical band width in loudness summation. *The Journal of the Acoustical Society of America*, vol. 29, no. 5, pp. 548–557.